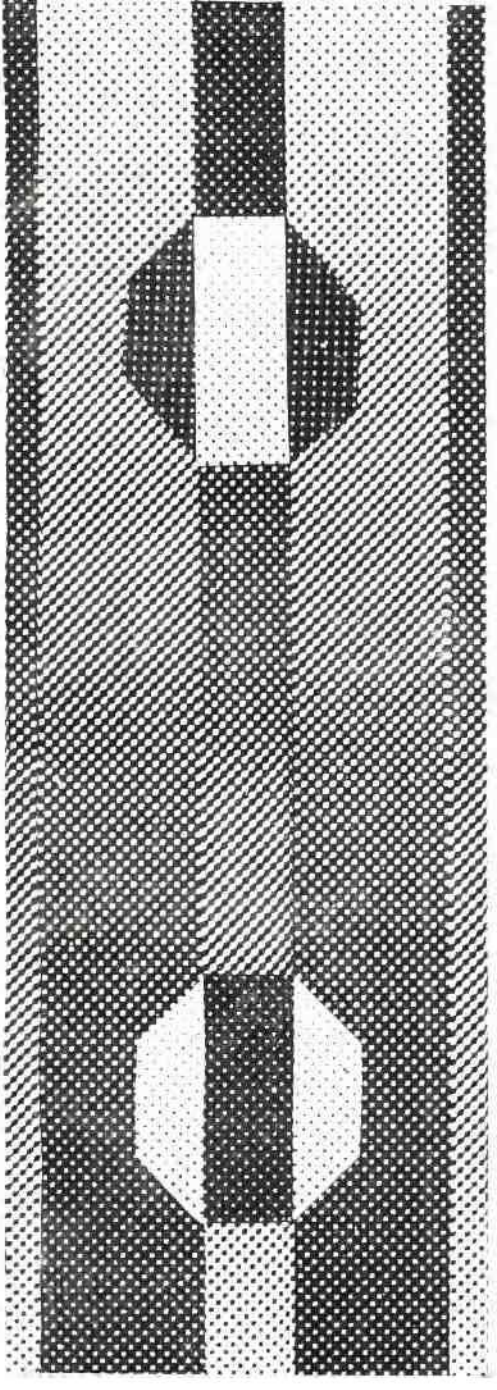


社会研究方法

VISIT…

社会研究方法

〔美〕艾尔·巴比 著
李银河 编译

四川人民出版社
一九八七年·成都

责任编辑：唐正宇
特约审稿：贾新民
封面设计：戴士和
封 面 画：施本铭
插　　图：孙　敏
装　　帧：盛寄萍

走向未来丛书

社会研究方法

〔美〕艾尔·巴比 著
李银河 编译

四川人民出版社出版（成都盐道街三号）
四川省新华书店发行
自贡新华印刷厂印刷

开本787×960mm 1/32 印张 14.625 插页 5 字数 199 千
1987年5月第一版 1987年5月第一次印刷
ISBN7—220—00170—3/C·56
书号：17118·167 印数：1—53,000

定价：2.28 元

Earl Babbie
Wadsworth Publishing Company
Belmont, California
A Division of Wadsworth, Inc.
1983年 第三版

社会研究方法

〔美〕艾尔·巴比 著 李银河 编译

蒋 琦 王小波 校

《走向未来》丛书顾问

（按姓氏笔划为序）

包遵信　严济慈　杜润生　张黎驷　陈一咨
陈翰伯　钟沛璋　侯外庐　钱三强

《走向未来》丛书编辑委员会

主　编：金观涛
副主编：陈越光　唐若昕
编　委：（按姓氏笔划为序）

丁学良　王小强　王岐山　王军衔
王晓鲁　王　焱　尹蓝天　乐秀成
朱嘉明　朱嘉豪　刘　东　刘青峰
严家其　何维凌　张　钢　阮芳赋
陈子伶　陈越光　易小冶　金观涛
秦晓鹰　贾新民　翁永曦　唐若昕
陶德荣　黄江南　董秀玉　樊洪业
戴士和

编者献辞

《走向未来》丛书和读者见面了。

她凝聚着我们的心血和期望。

我们期待她能够：展现当代自然科学和社会科学日新月异的面貌；反映人类认识和追求真理的曲折道路；记录这一代人对祖国命运和人类未来的思考。

我们的时代是不寻常的。二十世纪科学技术革命正在迅速而又深刻地改变着人类的社会生活和生存方式。人们迫切地感到，必须严肃认真地对待一个富有挑战性的、千变万化的未来。正是在这种历史关头，中华民族开始了自己悠久历史中又一次真正的复兴。

在艰苦而又富有生命力的改革道路上，我们坚定了马克思主义的信仰，理解了科学的价值，并逐步深化了对我们时代和民族的认识。今天，我们听

从祖国的召唤，热情地投身于实现社会主义现代化的伟大潮流。

马克思有一句名言："思想的闪电一旦真正射入这块没有触动过的人民园地，德国人就会解放成为人。"* 今天，照亮我们民族的思想闪电，就是马克思主义、科学精神和我们民族优秀传统的结合，以及由此开始的创新！

《走向未来》丛书力图从世界观高度把握当代科学的最新成就和特点，通过精选、咀嚼、消化了的各门学科的知识，使读者特别是青年读者能从整个人类文明曲折的发展和更迭中，理解中华民族的伟大贡献和历史地位，科学地认识世界发展的趋势，激发对祖国、对民族的热爱和责任感。

她特别注重于科学的思想方法和新兴的边缘学科的介绍和应用；把当前我国自然科学、社会科学，以及文学艺术方面创造性的成果，严肃地介绍给社会，推动自然科学与社会科学的结合。

《丛书》是个新的园地，她将自始至终贯彻严肃认真的学风和生动活泼的文风。

《走向未来》丛书，从她一开始就受到老一辈共产党人关怀，受到学术界前辈的热情支持。

---

* 《马克思恩格斯选集》，人民出版社，1976年版，第1卷，第16页。

约四百年前，弗兰西斯·培根在《伟大的复兴》一书序言中，曾经这样谈到书中描述的对象，他“希望人们不要把它看作一种意见，而要看作是一项事业，并相信我们在这里所做的不是为某一宗派或理论奠定基础，而是为人类的福祉和尊严……。”我们怀着真挚的感情，把这段话献给《丛书》的读者，希望广大读者关心她、批评她、帮助她。

让她成为我们共同的事业。

《走向未来》丛书编委会

一九八三年六月于北京

前　言

这是一本社会科学方法论的入门书。说它是入门书意思是，如果这书中的内容你全了解，并不意味着你已经可以搞出高水平的研究；但是如果对这书中的内容你还一无所知，却可以肯定地说，你还不懂得如何搞社会科学研究——你还没有入门哩。

很长时间以来，我一直想介绍一些西方的社会科学研究方法，原因部分地是由于对一位领导同志曾表达过的思想有同感，他说，中国不但在自然科学方面落后于西方，在社会科学方面也是落后的（大意如此）。落后就要向先进的学，学什么呢？首先想到的就是学些方法。方法是工具，西方人能用东方人也能用，就象一把螺丝刀一样。

方法论的书如汗牛充栋，为什么选中了这本

书？有必然性也有一点偶然性。书写得好，全面系统，深入浅出，有大量生动的例子，这些都属于必然性。偶然性是指两件小事。一件是在我做助教的匹兹堡大学社会学系本科生方法论的课上，一位在美国小有名气的方法论教授选这本书为课本；另一件是当我请我的导师（他是一位以方法论见长的英籍教授）推荐一本“最好的”方法论书时，他推荐了这本书。

这本书适于什么人读？我看大学本科生可以（虽然一般印象中中国的大学本科生比美国的水平高些），对社会科学研究感兴趣但未受过专门训练的人也可以读读这本书。

最后关于这种“编译”方式还想说几句。有人对“编译”颇有微辞，认为译书不全译会歪曲原著。无奈出版编辑给了篇幅限制。我的编译方法是一字一句读原文，把主要意思全放进去，删掉了一些例子。自信还是忠实于原著的，甚至使书比原著更精练了些。另外编是版权问题了。我在此郑重声明，这是一本译著，我只是译者，虽然翻译时未经过著者同意，但由于中国尚未加入国际版权组织，这样做还不能认为是不道德。等我国加入版权组织之后，该给他们交多少税，只好到时再说了。

别不赘述。向读者致以最美好的祝愿——愿你们喜欢这本书。

李银河

于美国匹兹堡1986年1月30日

目录

第一篇　总论 …… 1

第一章　人类的研究方式与科学 …… 3

一、两种真实性 …… 6

二、人类的初始研究方法 …… 7

三、研究者的失误 …… 10

四、社会科学的基础 …… 17

第二章　理论和研究 …… 23

一、社会科学理论之形成 …… 25

二、演绎理论之建立 …… 31

三、归纳理论之建立 …… 38

四、理论与研究的关系 …… 39

第三章　因果关系 …… 41

一、决定论与社会科学 …… 43

二、解释的个性与共性模式 …… 48

三、因果关系的标准 …… 50
四、度量与相关 …… 51
第二篇 研究之构成 …… 59
第四章 研究设计 …… 61
一、研究目的 …… 64
二、分析单位 …… 67
三、研究题目 …… 72
四、时间尺度 …… 74
五、研究动机 …… 76
六、研究设计 …… 79
第五章 概念之定义与度量 …… 85
一、现存事物之度量 …… 87
二、定义与研究目的 …… 93
三、度量的质量 …… 94
第六章 概念之具体化 …… 99
一、概念之具体化 …… 101
二、概念具体化范例 …… 108
三、提问规则 …… 110
第七章 抽样逻辑 …… 115
一、抽样的历史 …… 117
二、概率抽样的逻辑 …… 120
三、抽样的概念和术语 …… 122
四、概率抽样理论和抽样分布 …… 125
五、抽样设计之类型 …… 133

六、举例：在大学生中抽样 …… 136
七、多段类集抽样 …… 137
八、举例：在女教徒中抽样 …… 143
九、非概率抽样 …… 145

第三篇　观察方式 …… 151
第八章　实验 …… 152
一、适于采用实验方法的研究课题 …… 153
二、古典实验方法 …… 154
三、研究对象之选择 …… 158
四、实验设计的不断改进 …… 161
五、自然性质的实验 …… 168
六、实验方法之得失 …… 168
第九章　调查研究 …… 171
一、适于采用调查研究方法的研究课题 …… 173
二、问卷设计 …… 174
三、自填式问卷 …… 185
四、访问调查 …… 188
五、两种调查方法之比较 …… 196
六、调查研究方法之得失 …… 196
七、第二手资料的分析 …… 198
第十章　实地研究 …… 199
一、适于采用实地研究方法的研究课题 …… 202
二、观察者的各种角色 …… 203
三、实地研究的准备工作 …… 206

四、实地研究中的抽样 …… 207
五、提出问题 …… 209
六、记录观察内容 …… 209
七、数据处理 …… 210
八、数据分析 …… 212
九、得出结论 …… 214
十、实地研究范例 …… 216
十一、实地研究方法之得失 …… 217
第十一章 非接触性研究 …… 221
一、适于采用内容分析方法的研究课题 …… 224
二、内容分析中的抽样 …… 225
三、内容分析中的编录方法 …… 227
四、内容分析范例 …… 231
五、内容分析方法之得失 …… 233
六、现存统计资料分析 …… 233
七、历史-比较分析 …… 239
第十二章 评价研究 …… 245
一、适于采用评价研究方法的研究课题 …… 247
二、评价研究的设计 …… 248
三、某些有关的社会因素 …… 255
四、社会指标研究 …… 258
第四篇 数据分析 …… 263
第十三章 资料之数量化 …… 265
一、关于计算机 …… 267

二、数据处理术语 …… 270
三、编录 …… 270
四、编录表的设计 …… 274
五、编录与数据输入的几种选择 …… 277
六、数据之清理 …… 278
第十四章 基本分析方法 …… 281
一、单变量分析 …… 283
二、分组比较 …… 288
三、双变量分析 …… 289
四、多变量分析简介 …… 290
第十五章 指标、尺度和分类 …… 297
一、指标与尺度 …… 300
二、指标设计 …… 302
三、尺度设计 …… 311
四、分类 …… 315
第十六章 阐释模式 …… 317
一、阐释模式发展史 …… 319
二、阐释模式 …… 324
第十七章 社会统计学 …… 331
一、描述统计学 …… 333
二、推论统计学 …… 340
第十八章 复杂分析方法简介 …… 351
一、回归分析 …… 353
二、时间系列分析 …… 358
三、路径分析 …… 359

四、因素分析 …… 361
五、最小间隔分析 …… 363
六、方差分析 …… 364
七、对数线性分析 …… 368

第五篇 研究的社会背景 …… 370

第十九章 社会科学研究中的伦理和政治问题 …… 371
一、概论 …… 373
二、社会研究中的伦理问题 …… 374
三、两个关于伦理问题的争论 …… 379
四、社会研究中的政治问题 …… 386

第一篇 总 论

科学一词尽人皆知，然而究竟什么是科学则众说纷纭。对某些人来说，科学就是数学；对另外一些人来说，科学则是白大褂和实验室。科学当然不是这些东西的直接代名词，然而给科学下一个确切的定义亦并非易事。本书将把科学看作一种研究方法——一种认识和了解我们周围世界的手段。

当开始读这本书时你会发现，对于社会研究的科学方法你已经知道很不少了。事实上，在日常生活中，你一直在做科学研究。从这一意义上可以说，本书的目的在于帮助你完善那些已经掌握的研究方法，同时告诉你一些新的研究技巧。

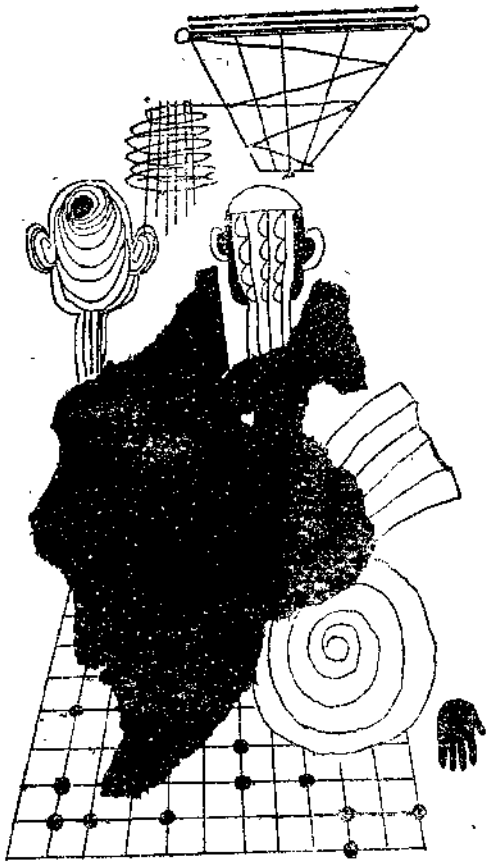
本书第一篇系全书讨论的基础。该篇综述了科学方法区别于其他了解事物方法的基本特征。在第

一章我们要讨论初始的研究方式，即人们在日常生活中所通常使用的方式。我们将讨论人们在研究周围事物的过程中怎样误入歧途，以及防止这类错误的方法。

第二章专门讨论社会科学的研究，社会科学理论的结构与形成，以及理论与研究的关系。本章的讨论适用于研究人类的社会行为。虽然对于人的研究有其特殊性，但一切科学的基本逻辑是相同的。

为发展概括性的理论，科学家致力于探索各种变量之间相互关系的模式。这些关系常常表现为因果关系。第三章就是讨论社会科学研究中因果关系的性质和逻辑。本章理论性的讨论是其余章节有关分析技术的讨论的基础。

总而言之，第一篇的宗旨在于提供一个背景，以便在此基础之上讨论更为具体的研究设计和实施。



第一章
人类的
研究方式与科学

本书的目的在于帮助你掌握了解事物的方法。让我们从几件你知道的事情谈起。你知道人为了生存必须呼吸，也知道月球的阴影部分是冷的，还知道中国人说的是中文。你是怎么知道的？因为有人这样告诉过你而你也就这样相信了。另外，有些事似乎太显而易见了。假如有人问，你怎么知道人为了生存必须呼吸，你会说“这人人都知道”。诚然，有许多事情是“人人都知道”的。例如，“人人都知道”地球是圆的；然而，几百年前“人人都知道”地球是平的。

许多你自认“知道”的事情不过是你赞同的或相信的罢了，很少是亲自经历或发现的。这并没有什么不对，达成某种一致的见解，正是人类知识的根本基础。然而，除间接了解事物之外，人们也可以通过直接经验得到知识。假如跳进冰冷的河水，

你马上会知道水是凉的，这是你通过自身感觉而了解到的。你第一次踏在荆棘上马上就知道疼痛而不需要有人事先告诉你。

1 两种真实性

由此可知，我们所生活的世界具有两种真实性。在你所了解的事物中，一部分具有“赞同性的真实性”，另一部分则具有“经验性的真实性”。前者是别人告诉你而你信以为真的事物；后者则是你通过直接经验所了解到的事物。问题在于二者看上去都具有真实性。

试举一例。你去出席一个高雅的宴会，发现有一种食物味道极佳，你吃了许多而且赞不绝口，可是，当主人告诉你这东西是炸蚯蚓时，你可能会顿时恶心起来并且吐了一地。这个例子说明，对食物的两种感觉都具有真实性，最初喜欢这食物的感觉和后来恶心的感觉都是真实的。然而，应当指出，你知道吃了蚯蚓而感到恶心完全是出于一种赞同，出于你赞同众人所说的蚯蚓是不能吃的。如果没有这种信念作祟，蚯蚓为什么会引起恶心呢？在那些没有这种信念的社会里，蚯蚓确是美味哩。因此，到底什么是

真实就成了一个微妙的问题。怎么才能知道什么是真实的呢？人们争论了几千年，科学就是从这些争论中产生的。

科学提供了一种把赞同的真实性与经验的真实性统一起来的方法。科学家对于那些未曾亲身经历过的事物运用某些标准去衡量。总的来说，一种说法必须同时具有逻辑的和经验的可信性才可接受。科学家根据什么接受了月亮阴影部分是冷的这一说法呢？第一，它符合月亮表面的热度是太阳照射而来的逻辑；第二，对月亮阴影部分的科学测量证实了这一预言。因此，科学家接受了未经亲历的事物的真实性——赞同性的真实性，当然是根据一些独特的标准来接受的。

2 人类的初始研究方法

人和许多动物都表现出预见的欲望。人们常常运用因果论和概率论的方法来预言未来。首先，人们一般认为将来的情况是由现在造成的：上大学会影响将来的收入；游过防鲨网会碰上鲨鱼；平时用功则考试会得好分数。其次，人们还知道因果关系实质上是符合概率的：得到某种效果的可能性在起

因出现时比起因不出现时要大，但并不是绝对的。在多数情况下用功就能考出好成绩，但并不能保证次次如此；游过防鲨网很危险，但并不是次次会碰上鲨鱼。

在初始的人类研究中，应当区别预言和理解。人们常常可以作出预言而并不理解为什么会如此，而且即便在不理解原因的情况下，人们也往往会按照自己的预想去行事。试举一例。某个赛马迷发现有一匹马总是赢，就有可能一直把赌注下在那匹马上，可是他并不知道也不想知道为什么总是那匹马赢。然而，如果人们能理解事物间的内部联系，以及某种模式一再出现的原因，而不是仅仅停留在对这些模式的观察和记忆上，就能够作出更为准确的预言。人类研究的目标就在于回答“什么”和“为什么”。

如前所述，我们对世界的认识只是部分地来源于直接研究或亲身经历，而更多的是从别人那儿得来的赞同性的知识。这种第二手知识对人类的研究既有利也有弊，其主要来源有二：传统和权威。

（1）传　统

每个人都会继承某一文化中认为天经地义的知识。例如，人们相信在春天播种玉米会得到上帝的

保佑，吃糖过多会损坏牙齿，手淫会导致双目失明。虽然我们有可能去亲自检验一下这些“真理”，但对于其中的绝大多数我们只是简单地接受了，因为这些事情是“人人都知道”的。

在这种意义上，传统对于人类研究不无好处。由于你接受了这些“人人都知道”的知识，研究的负担就大大减轻了。知识的积累性与继承性提供了其进一步发展的起跳点。所谓“站在巨人的肩膀上”就是站在前辈人的肩膀上。

同时，传统对于人类研究亦有害处。假如某人想对某些“人人都知道”的知识作出新鲜的与不同的理解，很可能被视为愚不可及甚至被送进精神病院。更加可能的是，人们根本不会产生对那些已知的及显而易见的事物作出不同理解的想法。

（2）权　威

尽管传统的力量很强大，新的知识还是层出不穷。在个人的研究之外，人的一生都在享用他人的新发现及对事物的理解。而对这些新知识的接受程度则往往取决于发现者的身份。流行病学家说流感会通过接吻传染似乎就比一个保姆说同样的话更可信些。

象传统一样，权威对于人类研究也有利有弊，

在某些事情上，特别是在那些有争论的问题上，信从受过特别训练的、有专业知识的、有职业证书的人的判断，会有利于我们对事物的研究。同时，某一领域权威的错误观点亦会严重阻碍研究工作的进行。生物学家在生物学领域中犯错误是不足为怪的。

此外，如果我们依赖某些专家权威对其专门领域之外的事物所发表的见解，也会妨碍我们对事物的正确理解。试想，让缺乏生物化学知识的政治或宗教领袖谈论大麻是一种危险的毒品会有什么效果。广告业滥用权威的例子更是屡见不鲜。比如，让体育明星讨论麦片的营养价值，让电影演员评论汽车的优劣等等。

由此可见，传统和权威都是研究事物的双刃剑。它们既为我们的研究提供出发点，又有可能把我们的研究引入歧途。

3 研究者的失误

除传统和权威所造成的潜在危险之外，研究者在研究过程中还会犯其他错误。这一节将论及研究过程中常见的失误及防止这些错误的科学手段。

（1）观察的不准确性

研究的关键在于观察。在研究“为什么”之前必须知道“是什么”。然而，对于千头万绪的日常事体，我们往往是粗心的和不自觉的观察者。例如，你能记得刚才与你说过话的那个人穿的是什么鞋吗？我们的观察有很大的偶然性因而常常是错的。一个美国游客在法国遭到几个陌生人的无礼对待便得出法国人粗暴无礼的结论，殊不知那些冒犯者却是一群德国游客。

与这种随意观察相反，科学的观察是一种有意识的活动。单是周密规划本身便能减少观察的误差。你可能不会记得某教师在课上穿的是什么样的衣服，但是如果你在课前就有意识地对观察和记录作出规划，观察的准确性便能大大提高。

在许多情况下，简单的或复杂的测量工具都有助于防止不准确的观察，同时在精确度上它们也大大超过了人的感觉所能达到的水平。例如，用彩色胶卷拍下那位老师的装束。

（2）以偏概全

当我们在周围的具体事物中寻找模式时，常把少量类似的情况当成普遍模式。这种以偏概全的倾

向会把研究引入歧途。

假定你是一名记者，被指派报道一次反核示威游行，并要求在两小时内交稿。为了了解人们参加示威的原因，你匆忙赶到现场采访。如果头两个接受采访的示威者恰恰给了你相同的原因，你便假定其余三千人也是出于同样的原因来参加示威的，这就犯了以偏概全的错误。

科学的研究用设计足够大的调查样本的方法来防止以偏概全的错误(见第七章)。此外，重复观察，看每次观察结果是否一致也可防止类似错误。由其他研究人员进行完全独立的重复观察则更为有效。

(3)观察的片面性

以偏概全的危险之一在于它会导致观察的片面性。人们一旦认为某种特殊模式是存在的，便会特别注意那些与此模式相符的事实而忽略与其不符的事实。假定有一次某人被一个犹太店主骗了，因而得出结论：犹太店主都不诚实。结果，他可能专门去注意其他犹太店主的不诚实行为，而忽视了诚实的犹太店主和不诚实的非犹太店主。有些人专门注意那些懒惰的黑人而忽略了勤奋的黑人和懒惰的白人。还有人专门注意那些非理智好冲动的女人，而忽略了情绪稳定的女人和情绪不稳定的男人。

为防止这类错误，一项研究设计通常要在事先确定调查的数量和种类，并在分析全部调查资料的基础上得出结论。科学的研究还采取另一种办法来防止观察的片面性，即同行们的相互监督和提醒，使人们注意那些容易忽略的或与自己的结论相反的事实。这是科学家互相提供的一种服务，也是科学事业本身的需要。

（4）信息推理

有时人们简直无法忽略某些与自己的结论相反的事实。例如，当他认为所有的犹太店主都不诚实时，一位犹太店主步行四英里归还了他丢失在柜台上的钱包。怎样看待这件事？人们常常另找理由来解释这个矛盾现象，如：也许这个店主并不真是犹太人，也许他是借送钱包之机来查看房址以便行劫。当科学观察与分析同原先的假设有出入时，人们往往寻找理由来解释这些意外的结果。

如果我们假设了解时事较多的人会较少感到政治异化。研究是在一群大学生中进行的，而结果是了解时事多少与政治异化感无关。我们开始另想原因：这是因为大学生一般都较关心时事，只有一般公众里才有真正不了解时事的人。如果到他们中间去调查，原先的假设就有可能成立。这就是所谓的

追溯假设，只要不就此止步，在科学上是完全可以接受的。推理本身并不能证实这一新假设，但末后的调查有可能证明它的准确性。因此，科学家经常进行信息推理，并在其后对事实本身进行重新调查。

（5）非逻辑推理

当观察与假设不符时人们常以“例外证明规律”来解释。其实例外只能引起我们对某一规律的注意。在任何逻辑体系中例外都不可证明与其矛盾的规律。统计学家所称“赌徒幻觉”便是日常生活中另一种非逻辑推理。它假定长时间赢后必输，长时间输后必赢。正是这种错误信念支配着许多手气不好的赌徒们彻夜地赌下去。虽然我们在日常生活中难免会陷入非逻辑推理，但科学家却能够自觉地和明确地运用逻辑体系以避免这种错误。

（6）理解中的自我介入

对一般性规律的寻求和理解并非单纯的知识活动，它直接影响着人们的个人生活，因而具备特别的心理意义。当人们失业或失去提升机会时，往往认为这是雇主想空出位置给自己的朋友，这一解释免除了人对自己能力和价值的怀疑。科学家作为常人也难以避免科学研究中的自我介入。设想某人由

于发现某种治癌方法而获得了诺贝尔奖金，而有人却写文章贬低他的方法，这时他很难做到完全客观地看待这种批评。遵从科学研究的规范可以避免过多的个人介入，而对批评意见持更为客观的态度。因此，自我介入虽然对某些科学家个人来说是个问题，但对科学整体却威胁不大。

（7）过早结束研究

以偏概全，观察的片面性，信息推理和非逻辑推理都会导致研究的过早结束。当一个反犹主义者说“我已经了解犹太人了，不要再用事实把我搅糊涂”时，他已经中止了对这一问题的研究。

过早结束研究的危险性是显而易见的。回顾人类知识发展史，我们发现，人类一直在改变着对已知事物的看法。从这个意义上说，研究是无止境的。科学是一个开放型的事业，结论总是被不断更新。纵然某个科学家认为某项研究已彻底结束，其他人却不以为然。甚至当整整一代科学家认为某一问题的研究已告结束时，后代科学家还会重新检验并改变那些旧的观点。

（8）对难以解释因素的神秘化

没有人能够通晓世间万物。不论我们多么聪明

和勤奋，还是有无数事件和情况是我们难以解释的。对于这些人类难以理解的事物，人们往往诉诸超自然的或神秘的原因。这种作法就是所谓“对难以解释因素的神秘化”。宗教或巫术对事件作出尽管是超自然的，然而却是合乎某种情理的解释。上述用语则不然，它只不过声称某些事件的起因终将超出人类理解力的范围。对于那些完全属于随机事件，这也许是对的。但接受这种不可知论，则可能导致研究的终止，不顾研究对象究竟是可知的还是不可知的。

防止对难以解释现象的神秘化就是要相信科学。科学认为一切事物都是可知的，或具有潜在的可知性。即使某位科学家认为某一现象超出了人类的理解力，其他科学家则可能设法去理解它。如果我公开宣称“爱”是不可理解的，也许反而会激励你寻找途径去理解“爱”这一现象。

4 社会科学的基础

科学被描述为逻辑一实证的，这意味着科学的两大支柱是：（1）逻辑或理性，（2）观察。换言之，对世界合乎科学的理解必须有意义并且同实际

观察相符。两者缺一不可。

粗略地说，科学的理论处理逻辑方面的问题；而研究处理观察方面的问题。科学的理论描述世界万物间的逻辑；而研究提供证明这些关系存在与否的手段。

（1）理论，而非哲学或信仰

社会科学理论所关心的问题在于“是怎样”而不是“应当怎样”。许多世纪以来，社会理论对二者不加区别。社会哲学家把“是什么”、“为什么”和“应当怎样”混在一起谈。然而，科学理论只涉及“是什么”和“为什么”，不能解决价值观念的争端。科学不能确定资本主义与社会主义孰优孰劣，除非在某些指标上有一个共同的尺度。只有首先对人类尊严和自由这类抽象概念达成一致的衡量标准，才能对究竟是资本主义还是社会主义能最大地满足人类尊严和自由作出科学的判断。而其结论除说明这一问题外并无其他广泛意义。在现实中，人们很难在价值观上取得一致，因此科学很难用于解决价值观的争端。况且人们对价值观的信念多是非理性的，这就使得以理性为特征的科学更加不便涉足。

（2）社会规律性

社会科学理论的最终目标在于确定社会生活规律的逻辑性和持续的模式。这个目标建立在一个基本的假设之上：生活是有规律可循而不是完全混乱或随机的。这个假设适用于所有的科学，但初看上去自然科学的研究对象似乎远比社会科学的研究对象更有规律性。投掷一重物，每次它都会往地上落；而一个选民却可能对某位候选人这次投赞成票下次投反对票。当温度上升到一定高度时冰总会融化；而一个看上去很诚实的人有时却会偷东西。这种情况真则真矣，却可能使我们因此忽略了社会现象中的高度规律性。

首先，大量的正式规范使得社会生活在相当程度上有章可循。例如，人要达到一定的年龄才有选举权。这种正式规定使社会行为规范化。其次，其他非正式社会规范造成了更多的规律性。例如，大学教授比非技术工人工资高，男人比女人工资高，白人比黑人工资高等等。

社会规范是确实存在的，而且社会科学家能够观察这些规范的影响。当规范随时间的推移而改变时，社会科学家可以观察并解释这些变化。当着社会学家指出按逻辑可以预见某种类型的人将以某种

方式行事时，这类人的行为应当是符合这一预见的逻辑基础的。

（3）总体，而非个别

对于研究社会规律性，应当明确指出，社会科学家研究的是社会的模式而不是个人的模式。一切有规律的模式都反映了由许多个人组成的总体的行动和情势。虽然社会科学家经常研究那些影响个人行为的因素，但个人本身很少成为社会科学的研究对象。因此，社会科学的理论是关于群体生活性质而不是关于个人的理论。

有时这种总体的规律性是令人惊异的。不妨以出生率为例。人们生育子女的个人原因是五花八门的：有些人出于满足自己父母的意愿；有些人是为了使自己的女性或男性臻于完美；有些则是为了维系婚姻关系；更有些是纯粹出于偶然。尽管个人的原因千差万别，年复一年，一个社会的总出生率却总是稳定的。如果今年的出生率是千分之十九点一，明年也会非常接近这个水平，尽管从长远来看这个数字会逐渐升高或降低。相反，如果有一个社会的出生率在连续五年中出现了千分之十九点一，三十五点六、七点八，二十八点九和十六点二的情况，人口学家便会茫然不知所措了。

（4）变量语言

关于由变量组成体系的思想较为复杂微妙，也许比喻可以使读者了解其中涵义。医生的对象是病人，他们行医是为了解除病人的痛苦；而医学研究人员则不同，他们的对象是疾病，病人只不过是疾病的载体。这并不是说医学科学家不关心人。他们研究疾病的最终目的就是为了解除病人的痛苦，但在实际研究中，病人只是作为被研究的疾病的载体。同样，社会科学研究的是变量及其属性，社会科学理论使用变量语言，而所涉及的人只是这些变量的载体。

属性是指客体的特征或特性，是用以描述客体的，如：女性的、东方的、异化的、保守的、不诚实的、聪明的等等；而变量则是这些属性的逻辑组合，例如：男性和女性是属性，性别则是由这两个属性构成的变量。职业这个变量由农民、教授、司机等属性组成。社会阶级这个变量则由上层、中层、下层这些属性构成。

变量与属性的关系是描述与解释客观事物的关键环节。例如，通过观察一个班级男性和女性这两个属性出现的频率，我们就可以用性别这个变量来描述这一班学生："这个班百分之六十是男生，百

分之四十是女生。”失业率可以被看作是用就业与失业这两个属性对变量就业状况的描述。

变量和属性的关系在用于解释事物时显得更加复杂。试举一例，其中包括教育程度和种族歧视程度这两个变量，每个变量采用两个属性。变量教育程度由受过教育与未受过教育两个属性组成；种族歧视程度由歧视与不歧视两个属性组成。假定频率分布如下表：

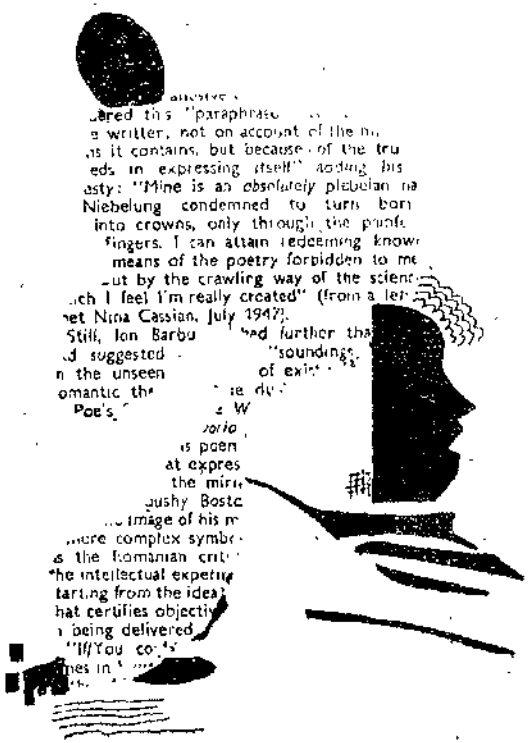
	受过教育	未受过教育
歧 视	3	9
不歧视	7	1

该表显示了教育程度和种族歧视程度这两个变量之间的关系。这一关系是通过两个变量各自属性的组合来描述的。占主导地位的有两组：（1）受过教育的并不持种族歧视态度的；（2）未受教育的并持种族歧视态度的。

理论描述变量之间的逻辑关系，它常常表现为因果关系。一个人在某一变量上的属性会导致他在另一变量上的某种属性，如：受过教育与未受教育分别导致不持种族歧视态度和持种族歧视态度。在这里，教育程度是自变量，种族歧视程度是因变量。在另一情况下，在本人教育程度与父母教育程度这

两个变量的关系中，本人教育程度又可以变成因变量，而父母教育程度则是自变量。

应当注意的是，理论在这里仅仅运用了两个变量：教育程度和种族歧视程度，而没有涉及人本身。如前所述，人是这两个变量的载体，只有通过观察人才能看到这两个变量间的关系。然而，归根结底，理论是由变量语言组成的，它旨在描述不同变量及其不同属性之间可能存在的一种逻辑的相关关系。



第二章
理论和研究

1 社会科学理论之形成

（1）科学的传统模式

科学的传统模式具有三个基本要素，它们按顺序是：理论、实施化和观察。

〔理论〕根据科学的传统模式，科学研究起始于科学家对现实世界某些方面的兴趣。例如，就整个社会而言，哪些重大事件会危及婚姻和家庭的稳定性呢？一个可能的答案是大规模的社会危机，战争、干旱、洪水、大量失业或类似的其他社会问题。

山姆·斯托芬(Sam Stouffer)，研究大萧条时的婚姻家庭状况便持这一观点。他认为大规模的危机影响了家庭的稳定性是有其逻辑根据的：首先，当

危机出现时，个人与家庭的日常生活必须作出相应的调整。由于家庭的稳定性依赖于常规和习惯，危机的出现便破坏了这一模式。其次，在发生某些危机如自然灾害时，政府和其他机构往往会介入而取代了部分家庭功能。当家庭经济难以维持时，政府通过救济取代了家庭的经济功能；而养家活口的传统功能的丧失则必然对家庭稳定性产生某种影响。虽然斯托芬的理论分析已大致阐明了各种变量之间的关系，但仅靠推理并不能对问题作出肯定的回答，还需要用观察来检验理论。

〔实施化〕斯托芬曾假设大萧条造成了冲动型婚姻的增加。问题在于如何识别这类婚姻以及从何着手研究。科学的传统模式的第二个步骤就是解决这类问题的。

实施化是对测量和辨认你所要观察的变量的具体步骤、程序的设计。斯托芬是通过分析官方婚姻统计资料来识别冲动型婚姻的。他采用了下列指标：（1）在新婚夫妇各自社区之外的婚姻；（2）由行政官而不是牧师主持的婚姻，（3）婚后五年内，尤其是在未生育子女情况下的离婚或分居，以及这些指标是否在大萧条时期有所上升。就这样，他制定出一个明确的计划：怎样去寻找问题的答案以及如何辨认和解释这些答案。

〔观察〕科学的传统模式的最后一个步骤是实际观察。观察方法包括实验、采访或实地观察。斯托芬在其研究中则是去搜集公开发表的有关统计资料。对于每一组统计资料，他都根据理论对可能出现的模式作出假设。例如他假设，如果大萧条促使被迫结婚的情况有所增加，那么在那段时间内由于婚前怀孕而造成的婚姻应有较高比例。该比例在1930年和1931年的急剧上升证实了他的理论预测。

图2—1为科学研究的传统模式之图解。从中可以看到，研究者首先对某一事物或有关该事物的资料产生兴趣，而后发展出一种理论上的理解，根据

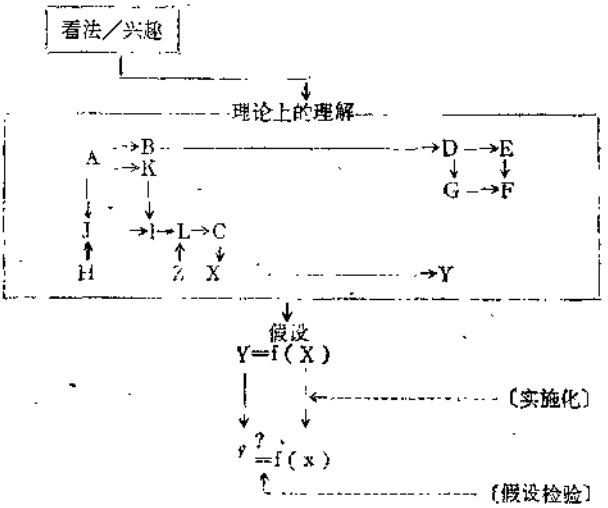


图2—1　科学研究的传统模式

这些理论上的考虑又产生了假设，即，如果理论成立的话，事情应该是怎么样。Y＝f（X）意为：Y（如被迫婚姻）是X（如大萧条）的函数（以某种方式受X的影响）。在这一层次，X和Y只有总体意义而无特别意义。在实施化过程中，总体概念被转换成具体的指标和程序。如小写x是大写X的具体指标。实施化过程旨在形成一个可被检验的假设。观察的目的在于发现，它只是假设检验过程的一部分。

（2）两类逻辑体系

上述科学的传统模式所使用的是演绎逻辑，本节将探讨演绎逻辑在社会科学研究中的应用并将它同归纳逻辑作一比较，科学哲学家比佛里奇(W.I.B Beveridge）对这两类逻辑体系做了如下的描述："逻辑学家将归纳推理（从特殊事件到一般原则，从事实到理论）与演绎推理（从一般到特殊，把理论运用于特殊事件）加以区别。归纳是从对数据的观察开始并从中发展出能够解释观察对象之间关系的概括性理论；而演绎则是从某些一般性规律出发并将其运用于特殊事件。"

演绎逻辑的古典例证是人所熟知的三段论："凡人都难免一死；S是人；因此S终有一死。"

如果运用归纳逻辑，则由S和其他所有被观察的人的死之得到一个暂时的结论：凡是人都终有一死。

图2—2比较演绎与归纳这种方法。我们想了解用于准备考试的时间与考试成绩之间的关系。运用演绎法是从检验事物的逻辑开始，考试分数反映了

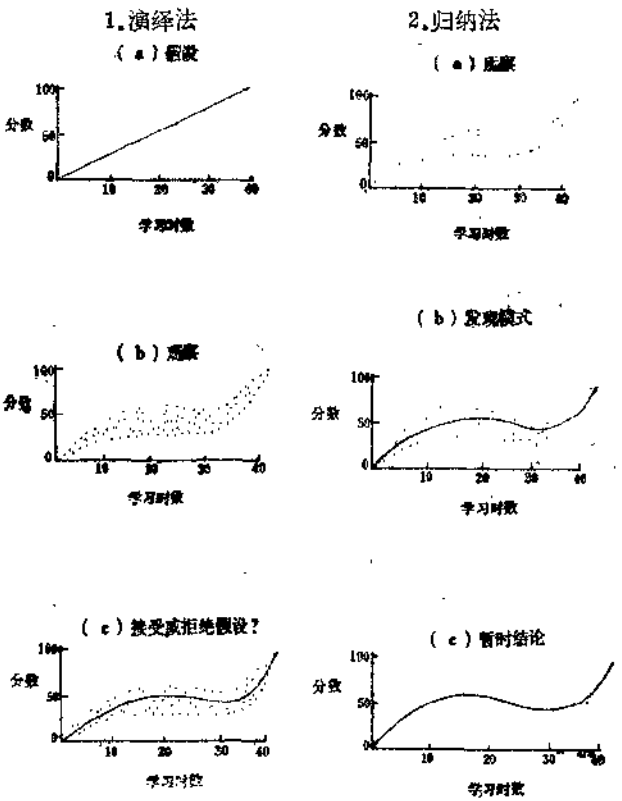


图2—2　演绎法与归纳法

学生记忆和使用信息的能力，而这两种能力都可以通过考试前对信息的了解加以提高。由此我们得出一个假设，即复习时数与考试分数有正相关关系〔图2—2第一部分（a）〕。下一步是进行与假设有关的观察〔第一部分（b）〕。最后将假设与观察进行比较〔第一部分（c）〕。由于观察与期望不可能毫无出入，我们必须判断二者的接近程度能否使假设成立。

同一问题运用归纳法则应从观察开始，即收集有关学习时数与考试分数二者关系的资料〔第二部分（a）〕然后找出对观察资料最具概括性的模式〔第二部分（b）曲线〕。最后，根据归纳法对两变量之间关系的模式作出暂时性结论，〔第二部分（c）〕。结论之所以是暂时性的，是由于我们所作的观察仅仅是模式的来源而不是模式的检验。

在实践中，理论和研究处于演绎与归纳的无限循环之中，华莱士（Walter Wallace)用一个环形来表示这个过程（图2—3）。

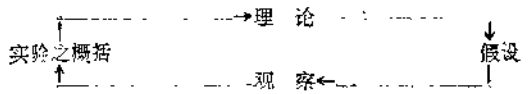


图2—3 科学环

根据华莱士的模式，由理论而产生假设，由假

设而导致观察，由观察而得出概括性结论，而这些概括性结论则改善了原有的理论。这些被改善的理论又导致被改善的假设及一整套新的观察，新的概括和被进一步改善的理论。如此周而复始，无限循环。这个模式的最大优点在于没有始点也没有终点，研究工作可以从任何一点开始。

2 演绎理论之建立

演绎理论的建立及假设检验包括哪些内容？下面先介绍一些有关演绎理论建立的术语，然后讨论如何建立演绎理论。

（1）与建立理论有关的术语

〔真实性〕关于真实性哲学家已经辩论了几千年，至今没有得到满意的答案。一个基本问题是究竟有没有独立于我们的经验而存在的事物。你手中拿着的这本书是实在的还是你的意识的产物？不错，你的意识告诉你双手感觉到书的存在，可是你的感觉也欺骗过你。由于不能最终证明是否真正有独立于我们的经验而存在的事物，我们做事时只好假定它是存在的。

〔客观与主观〕另一种方法是用客观和主观来解释真实性问题。我们承认某些事属于态度、意见和观点的范畴，如评价贝多芬和莫扎特谁是更出色的作曲家就属于主观的范畴，各人凭自己的经验作出判断。但我们感到手里这本书的存在则属于客观的范畴：它独立于我们的经验而存在。虽然"客观"的定义是"独立于意识之外"，但我们对客观存在的感知却是通过意识进行的。

〔观察〕在上述讨论中，对客观事物的"经验"系指人类感观的活动，在社会研究中一般限于视、听及接触。观察一词则泛指通过这些手段所进行的收集信息的活动。

〔事实〕虽然对于哲学家说来事实这一概念象真实性一样复杂，但运用于社会科学研究时，它一般是指被观察到的现象。例如：在1960年的总统竞选中肯尼迪战胜了尼克松便是一个事实。

〔规律〕卡普兰(Abiaham Kaplan)定义规律为对某类事实的普遍概括，如万有引力定律。然而规律必须是某类事实普遍的而不是偶然的模式。卡普兰举例指出，从1920到1960年每次美国总统竞选中都是名字最长的那位候选人获胜，这是事实但不是规律。它只是偶合，而不能说明在名字长度与政治成功之间存在什么必然的联系。规律有时又称为原

理，它们只能被发现而不能被创造。规律本身只是对事物的概括而不能作任何解释。理论的功能才是解释，而理论才是被创造出来的。

〔理论〕理论是对与生活某一方面有关的事实与规律的系统性解释。特纳（Jonathon Turner）提出理论至少包括四要素：概念、变量、陈述和形式。

〔概念〕特纳把概念称为“建筑理论大厦的砖石。”概念是对研究范围内各类现象的抽象。例如与青少年犯罪理论有关的概念包括“青少年”“犯罪”、“伙伴”以及“社会阶级”、“种族”等等。由于不同的理论对研究对象会作出不同的解释，因而使用不同的概念。如对于青少年犯罪，马克思主义的理论将强调同社会阶级及经济地位有关的概念。

〔变量〕理论不仅确定与其有关的概念，而且对这些概念加以定义，并将其进一步具体化为变量。如犯罪被定义为违反法律并进一步具体划分为轻罪和重罪。社会阶级可具体化为由下列属性构成的变量：工人阶级、中产阶级和上层阶级。

〔陈述〕理论由几类陈述组成。前述规律或原理是其中一种。“公理”是被视为千真万确的断言，它是理论的基础。“命题”是建立在公理之间相互逻辑关系基础上的有关概念之间关系的结论。

在青少年犯罪理论中，我们从公理“人人都希望得到物质享受”和“上层阶级比工人阶级具有较多的合法获得物质享受的能力”出发，得出“工人阶级子弟比上层子弟更有可能触犯法律以获取物质享受”的结论。这一结论便构成了一个命题。“假设”则是由命题引伸出发的对经验事实的具体期望。在上例中，理论可包含这一假设：“工人阶级青少年比上层阶级青少年犯罪率高。”而这个假设则可通过研究来加以检验。

〔形式〕构成理论的陈述必须被系统地组织在一个逻辑结构中。理论可用多种形式表达但必须符合一定的标准。它应当一目了然对问题给出一个“直观的解释”。

〔基本模式〕基本模式是组织我们对事物看法的基本理论框架。它不一定对重大问题作出直接回答，但为我们指出寻找答案的途径。目前，在社会科学研究中有三个占统治地位的基本模式：互动论，功能主义和冲突理论。

互动论把社会生活看作是个人间相互作用的过程。如用这一基本模式研究青少年犯罪问题，就会把研究重点放在伙伴的互动上。例如，这些年轻人如何通过观察朋友的眼神来获得认可？青少年团伙怎样吸收及教唆新成员？以及家长如何为子女提供

榜样等等。

功能论或社会系统论注重社会生活的组织结构。社会有哪些组成部分，它们怎样相互联系？从这一模式出发会特别注意广泛的经济状况对犯罪率的影响，如，高失业率时期犯罪率是否上升？现代社会日益增加的复杂性是否使年轻人难以辨别是非？造成社会失去控制及法律秩序解体的原因将会受到重视。

冲突理论把社会生活描述为相互竞争的个人与群体间的斗争。例如马克思的阶级斗争学说强调有产者与无产者之间的斗争。运用这一模式，青少年犯罪或许被视为工人阶级对统治者压迫所作出的反应，或青年人争取社会地位的一个途径。

这三种基本模式并不绝对相互排斥，但它们确实是从不同的侧面来认识社会生活的。

（2）起　步

建立演绎理论的第一步是选题。题目涉及范围可宽可窄但必须是你有兴趣的。一旦选定题目，即应着手开列出对该选题已知的情况和已有的想法。其中部分是你自己的观察和看法，部分是他人做过的研究。然后即应对与选题有关的主要概念加以定义。在这一阶段，你会发现一些在前辈学者中取得

一致意见的看法。如，一般说来老年人比年轻人保守；上层人士比工人阶级在经济上更倾向于保守、在政治上更倾向于自由化等。这些已知的模式对于你创造自己的理论会有很大的帮助。

（3）建立理论

建立理论的步骤并不是一成不变的，但下列要素可使之更加条理化：

①明确选题。

②明确理论的应用范围。如这一理论适用于全体人类的社会生活，还是仅仅适用于美国人，甚或仅仅适用于年轻人。

③确定主要的概念和变量并使之具体化。

④搜集和整理有关这些变量之间关系的已知命题。

⑤从这些命题出发对要研究的问题做逻辑推理。

（4）演绎理论范例

1965年夏天洛杉矶市郊华兹黑人居住区发生暴乱，在将近一周的时间内造成了三十五人死亡和二亿美元的损失。暴乱的参加者是什么人？他们为什么参加暴乱？社会学家兰斯福德（H.E.Ransford）提

供了一个答案。

他首先研究了有关极端政治行为的理论文献，发现过去学者们曾指出，社会隔离和权力丧失感往往与政治暴力相联系。根据华兹暴乱的情况，兰斯福德发现有理由认为这两个变量是暴乱的原因。他从当地黑人居民中选了一个样本来进行调查。关于社会隔离程度，他询问研究对象在工作地点、居住区域、社会团体及其他场合同白人接触的情况。关于权力丧失感和异化感则用另一些问题来测量。如他给出一些断言，诸如“世界掌握在少数有权力的人手中，小人物是无能为力的”，然后要求调查对象回答是否同意这个说法。根据调查对象的回答确定他们在社会隔离和权力丧失感方面的程度。由于人们害怕遭刑事起诉而不敢承认是否参加过暴乱，兰斯福德便向他们询问两个问题：（1）“为争取黑人权力你是否愿意使用暴力？”约四分之一的调查对象给了肯定的回答。（2）问他们是否已经这样做了，只有百分之五说做过。如果兰斯福德的理论假设是正确的，他应当发现社会隔离和权力丧失感严重的人更倾向于表示愿意使用暴力以及承认已经这样做过。结果确实如此。他的理论期望是有意义的，且为事实所证实。

科学中有两大重大因素，逻辑的完整和实验的

证实。单有逻辑是不够的，因为初看符合逻辑的推断并不见得符合事实；另一方面，单凭观察和事实也是不够的，因为它们不是科学的结论，不能提供对事物的理解。

3 归纳理论之建立

社会科学家创造理论也常常从观察社会生活的某个方面开始，以求发现具有普遍意义的模式。下面通过一个实例来说明归纳理论的建立过程。

为什么有些人比其他人更加笃信宗教？格洛克（Charles Glock）及其同事们1967年进行了一次大规模的调查。调查结果表明，有四个变量与宗教信仰程度关系最大：

①女人比男人宗教信仰强烈

②老人比青年宗教信仰强烈

③穷人比富人宗教信仰强烈

④单身者比有家庭的人宗教信仰强烈

这些差异可视为规律，但规律只示其然，未示其所以然。那么，为什么这四个变量影响了宗教信仰程度呢？经过数年研究他们发现了一个模式：每一种同笃信宗教相关的特征其价值在美国社会中都

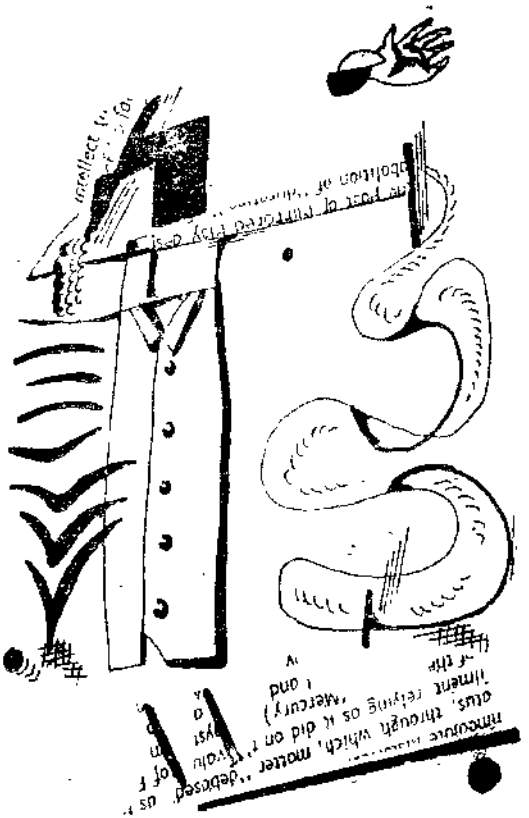
是被贬低的。如，女人比男人地位低，老年人比青年地位低，穷人比富人地位低，单身者比有家庭的人地位低。他们由这一模式得出结论："那些在世俗生活中被最大程度地剥夺了满足感和成功感的人求助于教堂以获得慰籍与补偿。"这就是所谓"宗教的社会剥夺理论。"

4 理论与研究的关系

本章介绍了在社会科学研究中理论与研究的关系。在演绎模型中，研究用以检验理论；在归纳模型中，则是通过分析研究资料来发展理论。在实际的社会科学研究中情况要复杂得多。有时理论仅仅给经验分析提供一个背景；有时实验数据也仅仅被引用来支持某一理论观点。在这两种情况下，都不存在以发展新解释为目的的理论—研究相互作用。有些研究甚至根本不涉及任何理论。

社会科学家在实际工作中是如何处理理论与研究的关系呢？威尔斯与皮可（Richaid H·Wills, J.Steven Picou）对1936—1978年《美国社会学评论》的内容分析研究表示，社会学家在研究中运用传统的科学研究模式——假设检验模式——的比

例有了显著的增长，由1936—1949年的16%，增加到1950—1964年的36%，直到1 65—1978年的54%。本章旨在说明，对于社会科学研究，并没有一份简单的“烹调指南”，但归根结底，科学依赖于两大支柱：逻辑和观察。



第三章
因果关系

1 决定论与社会科学

本书在与人类行为的自由意志相反的意义上讨论决定论的观点。一个基本的问题是：人的行为究竟是个人意志的产物还是那些人不能控制甚至根本未觉察到的力量和因素的产物。只有搞清了这一问题，才能进一步探讨因果关系在社会科学研究中的地位。

（1）物理学中的因果关系

让我们来做一个实验。拿起一枝铅笔，只要一松手，它就掉下去了。为什么？物理学家会说铅笔下落的原因在于地球引力。假如铅笔不想往下落会怎样？它仍会下落；假如铅笔很愿意往下落又会怎样？还是一样。极而言之，假如铅笔对自己说：

“我相信几铅笔都具有战胜地球引力的意志力。我要悬在空中。”事情又会怎样呢？还是一样。很难想象铅笔的意志会对万有引力产生丝毫影响。

地球引力一例说明了人们对因果决定论的常识性了解，它限制了人们行动的自由意志，正如小孩子都知道不管他们多么想飞还是飞不起来。物理学教给了我们严峻的一课：人的意志是受到某些决定性的限制。

（2）寻找引起偏见的原因

要寻找引起种族偏见的原因，令人满意的答案可能包括经济上的竞争、宗教意识、政治观点、童年经历、受教育程度等等。其基本逻辑是：人是否有偏见不是出于自身的选择，而是被某种力量推上这条或那条道路的，诸如童年经历、世代沿袭的宗教信仰或类似的不能控制甚至难以觉察的因素。

（3）原因之原因

然而人们可以这样来反驳上述推理：许多事情还是由个人选择的，这些选择反过来又决定了这些个人是否有偏见。例如，如果研究证明某人没有偏见是由于他所受到的教育，那么不是他自己选择去受教育的吗？这是否意味着他本人就是他无偏见的

原因呢?

这种观点的问题在于原因背后还另有原因。某些特殊原因如他的家庭世代都上大学因而他也想上，或他家世代都没有大学生而他想当第一代大学生。这两种因素都迫使他有上大学的强烈愿望，这种愿望迫使他上了大学，而上了大学又迫使他变得没有偏见。无论在这个过程的每一特定环节有什么原因，一个原因背后总是另有原因。上述讨论的含义在于人的偏见的原因可以顺着一条又长又复杂的因果链一直追寻下去，而这一系列的原因则对人的行为作出了解释。

再回到铅笔下落的例子。铅笔下落是由于引力的作用，同铅笔的意志毫无关系。这同偏见问题似乎不可同日而语。然而，当着我们进行解释性社会科学研究时——例如当我们去追寻偏见的原因时——我们所采用的人类行为模式则假定人并不比铅笔具有更多的个人选择自由。

（4）决定论带来的烦恼

人类行为的决定论模式是一个令人不快的模式。它令人不快的原因在于，假如人所做的一切都是人不能控制的力量的产物，那么人对自己所取得的一切成就也就无功劳可言了。从我们的文化价值

和信仰出发，相信了这一点生活就会变得索然无味。其次，人类行为的决定论模式使人感到好象掉进陷井而束手无策。尽管我们已经习惯于人不能飞这一事实，但承认每一行动都完全取决于象万有引力那样强大无情的力量还是相当困难的。此外，我们会忧虑接受了决定论模式我们将怎样去行动。试想当人相信自己的行为完全不受自己意志支配时还会不会再关心任何事情？人会不再想为改善周围世界有所作为，或者索性躲在角落里什么也不做。进一步想，人又怎么能够改变任何事物？显然不能，除非外界因素强制人去做。具有讽刺意味的是，人既然相信自己的行为不受个人意志支配，又何必去担心自己的行动？

（5）努比·戈德伯格(Rube Goldberg)的生态体系

美国漫画家努比·戈德伯格几年前的一幅作品现在成为生态学领域关于生态体系的现代观念的示意图。戈德伯格的漫画是一条环环相扣的事件长链（如图3—1）。一个油漆匠无意中撞了一下梯子打翻了梯子顶上的油漆，油漆溅到一只猫身上，它的尖叫把一个母亲吓了一跳，一松手婴儿车滑下去撞上一张饭桌，打翻了桌上的牛奶，撒在地上的牛奶

图3—1 努比·戈德伯格的生态体系

又滑倒了一个人，他手中的球飞了出去……这个事件链永无完结。

生物学家加里特·哈丁（Garrett Hardin）曾指出：生态学的第一定律是“人不可能只做一件事”。人所作的每一件事都有许多后果，宇宙中发生的每一件事都是无穷无尽的相互关联的事件链中

的一环，正如那个油漆匠撞了一下梯子所引起的一系列事件一样。在生态系统中，每个事件都与其他事件相关联，没有一件事是独立存在的。

注意图3—1是如何勾划出人类行为的决定性系统的。母亲为什么会丢掉了手中的婴儿车？因为猫的尖叫。可是猫又为什么要尖叫呢？仔细看一下图解，原来是母亲使得它叫起来的——由于丢下婴儿车所引起的一系列事件而使得猫儿尖叫起来。在生态体系中，任何因素都既是因也是果。两者都与事实相符，都是真实的。这就是人类行为的决定性模式所开的一个大玩笑；人在成为这个系统的结果时不能不同时成为其原因。

2 解释的个性与共性模式

社会科学运用两种模式来解释人类行为。前边讨论的多重原因决定某一特殊行为的模式称为解释的个性模式。这一模式的目的在于通过列举大量的独特原因去解释某一行动。传统的历史学家倾向于采用这种模式，如，他们列举一切导致法国革命，或美国决定参加第二次世界大战的特殊原因。大多数人都采用这种个性模式理解周围的事物。

科学家，包括社会科学家却常常运用另一种模式，称为解释的共性模式。这种模式不是去列举导致某一特殊行动或事件的全部原因，而是有意识地寻求可以解释行为或事件的一般类型的那些最为重要的因素。假定我们想了解人们在1980年美国总统大选中的投票原因。每个人都可能用许多原因来解释为什么投卡特或里根的票。假定某人给出99种选卡特的原因，我们就会认为这是对该人投票原因相当完备的解释。如果另一个人也提出这同样的99种原因，我们就可以相当有把握地预言那个人也会投卡特的票。这是解释的个性模式的作法。解释的共性模式的作法是从众多的因素中提炼出相对较少的因素，这些因素可以部分地解释许多人或所有人的投票行为。如政治倾向——自由或保守——可以成为决定全体选民投票行为的具有普遍性的重要因素。大多数具有自由派倾向的人可能赞成卡特，而大多数具有保守派倾向的人可能赞成里根。当然，这一单一的因素并不能对所有的选举行为作出全面的解释。解释的共性模式旨在用最少的原因变量去最大限度地解释因果关系的一般模式。解释的共性模式于是不可避免地采用概率的方法来解释因果关系。

3 因果关系的标准

保尔·拉扎斯菲尔德（Paul Lazarsfeld）为因果关系提出三项标准。

因果关系的第一个标准是在时间上原因先于结果。然而在分析社会科学数据资料时，两个变量的时间顺序往往并不明显。即使有时间顺序，例外也很常见。例如，我们一般假定父母的教育程度决定子女的教育程度。然而有些父母重返学校可能是因为他们的子女接受了高等教育。

因果关系的第二个标准是两个变量之间存在着实验的相关关系。然而在社会科学研究中，很难发现完美的相关关系。多数保守派成员选里很，但也有些保守派成员不选里根。因此只能规定实验的相关关系强大到什么程度则构成因果关系。

因果关系的第三个标准是两个变量之间的实验相关关系不能由于第三个变量的存在而消失。如，某人左膝一痛天就下雨，这并不意味着其关节对天气有任何影响。相对湿度这个第三变量才是导致膝痛与下雨的原因。

（1）必要原因与充分原因

在概率模式中应当区别两类原因：必要原因和充分原因。必要原因是指为某种结果的出现而必备的条件。例如必须是女人才能怀孕，虽然并非所有的女人都怀孕。充分原因是指只要其出现结果一定会出现的条件。参军是得到一套军装的充分原因，虽然通过其他途径也可以得到军装。

发现必要且充分的原因是研究的最高理想。可惜我们很少能找到必要且充分的原因。在实际中也很少有完美的必要原因和完美的充分原因。就社会科学而言，无论必要或充分原因都可构成因果关系。

4 度量与相关

（1）传统的演绎模式

传统的科学方法包括如下步骤：

①理论之形成

②理论性假设之形成

③概念之具体化

④实验数据之收集

⑤假设之实验检定

〔理论之形成〕科学家从现实世界中引起其兴趣的问题发展出抽象的演绎理论。假设他对越轨行为感兴趣，便在现有的社会学理论的基础上发展出一个越轨行为的理论，其中包括有关越轨行为起因的一系列概念。

〔理论性假设之形成〕在关于越轨行为的总体理论的基础上，科学家发展出与构成该理论的各种概念相关联的假设。如假设青少年犯罪是监督的函数：当着监督加强时，青少年犯罪便减少。

〔概念的具体化〕下一步便是将体现理论性概念的实验指标具体化。概念可以抽象甚至不太明确，实验指标却必须既精确又具体。如上例，概念的具体化包括下列内容："青少年"这一概念可规定为18岁以下；"犯罪"可具体规定为因犯罪行为而被捕；"监督"可规定为家里有不工作的成人。概念具体化的结果是把理论性假设转化为实验性假设。在同一例子中，实验性假设为：在18岁以下的青少年中，那些家里有不工作成人的比那些家里没有不工作成人的将较少因犯罪行为而被逮捕。

〔实验数据之收集〕根据具体化的理论概念，研究者收集与实验指标有关的数据。还用这个例子，研究者可在18岁以下的青少年中进行调查，调

查问卷将问及被调查者家庭中是否有不工作的成人以及是否因犯罪被捕过等等。

〔假设之实验检定〕待数据收集完毕，最后一步是用统计学方法来检定假设。科学家通过分析确定那些家中有不工作成人的青少年是否比那些家中没有不工作成人的青少年较少因犯罪行为而被逮捕。对实验性假设的肯定或否定于是被用来接受或拒绝理论性假设。

在运用这一模式时会遇到两个基本问题：第一，理论概念的具体化很难做到精确无误，因为概念是抽象的和一般的，每个具体的实验指标只能是近似的。例如"监督"这个一般性概念很难说可以由家中不工作的成人充分地体现出来。有不工作的成人并不一定能保证对青少年行为的监督，而某些没有不工作成人的家庭也可能作出其他安排来对青少年进行监督。第二，变量之间的实验相关关系几乎不可能是完美的。假如所有家中有不工作成人的青少年都从未因犯罪被捕过，而所有家中没有工作成人的青少年都因犯罪被捕过，我们于是可以下结论说假设被肯定了。反之，两组青少年的情况毫无差别，我们则可得出假设被否定的结论。但现实生活中这两种情况都不会发生。从实验上说，几乎所有的变量之间都有某种程度的相关。因而确定在何种

程度上可以接受或拒绝假设也是一种人为的规定。

概括地说，科学家是采用理论概念的近似指标去发现变量间的一部分相关关系。科学研究是一个旨在理解某些现象的永无止境的事业，为此目的，应当对现象之间的相关关系进行不断地度量和检验。

（2）指标的可置换性

保尔·拉扎斯菲尔德关于指标的可置换性的讨论为我们提供了理解度量与相关之间关系的概念工具，并部分地解决了上节提到的两个基本问题。他提出，任何概念都有数个可能的指标。如“监督”这一概念可以有好几个指标，写作x_1，x_2，x_3，等等。虽然有理由认为某些指标比其他指标好些，但从根本上说，它们是可以相互交换的。因此，当不知用哪项指标能更好地检验假设时，科学家采用所有的指标来进行一系列的检验。（见图3—2）

$$Y=f(X)$$

$$\downarrow \qquad \downarrow$$

$$y \overset{?}{=} f(x_1)$$

$$y \overset{?}{=} f(x_2)$$

$$y \overset{?}{=} f(x_3)$$

$$y \overset{?}{=} f(x_4)$$

$$y \overset{?}{=} f(x_6)$$

图3—2 指标的可置换性

（3）定点分析

上述关于指标的可置换性的讨论仅仅集中在一个概念上，事实上，所有的概念都具有这一性质。假如x可以通过多种方法度量，Y为什么不可以呢？例如我们想研究富有是否会产生保守主义。富有和保守这两个概念都可有多种度量方法。试举几例：

富裕程度（X）	保守主义（Y）
x_1 = 岁入	y_1 = 外交事务
x_2 = 存款	y_2 = 国内问题
x_3 = 财产/投资	y_3 = 经济问题
x_4 = 父母富裕程度	y_4 = 社会问题
x_5 = 生活方式	y_5 = 道德标准

鉴于有众多混杂的指标来度量富裕程度和保守主义，对这两个变量相关性的检验往往只能从散乱的指标开始（如图3—3第一部分所示）。由于两个变量都有许多可能的但尚未确定的指标，这就有必要先“固定”其中的一个指标，由此开始分析，这就是定点分析方法。如可用岁入（x_1）作为富裕程度的度量指标，检验它同保守主义各度量指标（y_1……y_5）之间的关系（如图3—3第二部分所

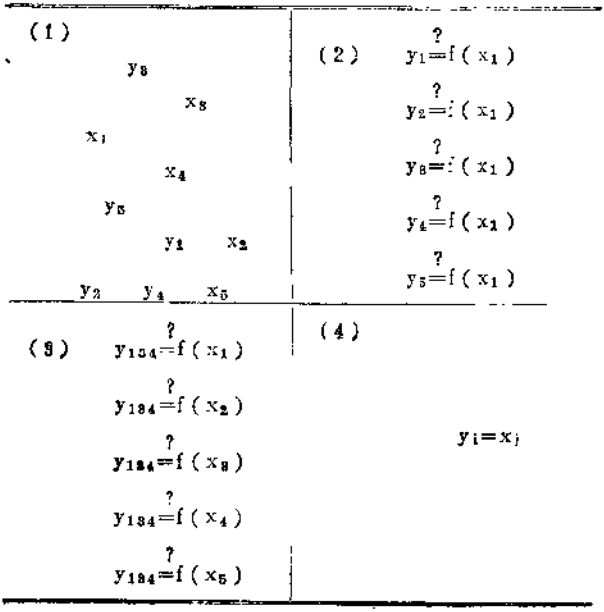


图3—3 定点分析

示）。假定分析表明富裕程度（由岁入度量的）同保守主义在外交事务、经济问题和社会问题(y_1，y_3和y_4）三个方面表现出强相关关系，我们可暂定用这三个指标的结合体来度量保守主义，用它来检验富裕程度的不同指标（x_1……x_5），（图3—3第三部分）其目的在于找出那些适合于研究保守主义的富裕程度指标。一旦确定了对富裕程度最有用的度量指标，便可用它来重新检验保守主义的各项

指标。最后，这个过程以确定可表明富裕程度与保守主义这两个变量之间关系的最佳度量指标而告结束，i和j是最终得到的度量指标（图3—3第四部分）。

（4）度量和相关举例

在宗教社会学中发生过一场关于宗教与种族偏见关系的辩论。查尔斯·格洛克(Charles Y. Glock)和罗德尼·斯达克（Rodney Stark）写了一本题为《基督教信仰与反犹主义》的书。书中提出实地调查数据表明，基督教教职人员中奉行正统派信仰的人比正统派信仰淡薄的人更加倾向于反犹主义。这一发现在教会中掀起轩然大波，其他研究者纷纷开展对该课题的研究。其中一项研究得到与格洛克和斯达克截然相反的结论。据研究报告称，正统派观念越强则偏见越少。可是如果仔细研究一下便会发现，该研究是采用传统的基督教教义“四海之内皆兄弟”及“爱你的邻人”等教义来度量信仰正统派观念的程度。那么接受这些教义的人比拒绝这些教义的人显得更少偏见当然是无足为怪的了。

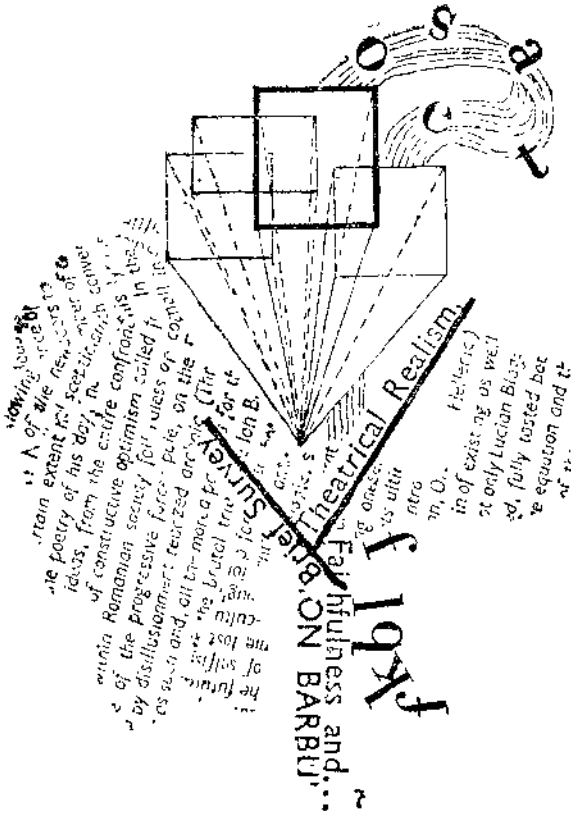
这样的研究结果通常会受到挑战，被认为是“概念混淆”：对正统派观念与偏见这两个概念所采用的指标实际上是度量同样的或类似的东西。如

果仅仅问正统派观念与偏见是怎样相互联系的而暂时不问它们是否相关，我们将得到如下结论：用格洛克—斯达克的指标所度量的正统派观念与偏见呈正相关关系；用遵从兄弟爱这类规范所度量的正统派观点与偏见呈负相关关系。两个结论在实验中都是正确的，但二者均未解决正统派偏仰与偏见是否相关这个大前提。

本章旨在介绍以非解释为目的的社会研究。虽然这不是研究的唯一目标，但我们常常希望能解释人们为什么如此这般地思考和行动，事情的原因和结果。对因果关系的介绍将有助于后面的技术性讨论。

第二篇　研究之构成

正确地提出问题往往比回答问题更加困难。一个正确表达的问题常常就是对问题本身的回答。本书第二篇即讨论“研究什么”这一问题。换言之，也就是正确地提出科学研究的问题和研究的构成。



第四章
研究设计

科学是一个“探索”的事业。任何探索都有多种途径。假如你想知道该不该买某种牌子的汽车，可以去买一辆试试其性能；可以去询问拥有这种车的人，或曾想买而结果并未买的人；可以去查报刊的分类广告栏看是否有许多这种牌子的旧车售价很低；可以看消费者杂志上对这种车的评价，或通过其他许多途径去了解。日常生活中如此，科学研究中也是一样。

研究设计是指对科学研究做出规划，即制定一个策略去探索某种事物。研究设计包括两个主要内容：第一，确定研究课题；第二，确定研究的最佳途径。

1 研究目的

社会研究可以有多种目的，择其要有三：探索性研究，描述性研究和解释性研究。

（1）探索性研究

许多社会研究是为着探索某一项课题，即对该课题进行初步了解。当研究者新接触一个课题或这个题目本身比较新鲜尚无人涉足时，研究便往往是探索性的。例如，纳税者普遍对政府不满而掀起了有组织的抗税运动。研究者可能想了解这一运动的广泛程度，社区支持这一运动的程度，这一运动是如何组织起来的，运动的积极分子是什么人等等。通过探索性研究，如向税务局官员核对数据，收集并研究运动文献以及参加会议并采访运动领导人等，即可对上述问题得到大略的答案。

探索性研究也适用于更持久一些的现象。如某学生对学校的宿舍规则不满意因而想改变它。他可以研究这个学校宿舍规则的历史，会见有关人员以了解制定这些规则的原因，与学生交谈了解他们对宿舍规则的大致看法等等。这最后一项工作不在于

对学生的意见做出描述，而是为今后更为周密的研究提供指导。

探索性研究主要有三个目的：第一，仅仅为了满足研究者了解某事物的好奇心和欲望；第二，探讨开展更为周密的研究的可能性；第三，发展可用于更为周密的研究的方法。作为第三点的一个例子是六十年代中期到后期，一所重点大学的研究人员所进行的对那个时期学生态度改变的程度、起因及后果的研究。在开展大规模的调查之前，他们预见到设计一个可以度量几种取向的问卷的困难。为此，他们开展了一项小规模的探索性研究，在50名可代表各种类型的学生之中进行开放式访问。访问过程中，被调查者运用自己的独特语言就有关他们对大学及社会的态度取向的问题做了深入的回答，从而提供了在进行大规模抽样调查时如何在结构性问卷中探测这种复杂性的方法。

探索性研究在社会科学研究中是很有价值的。但其缺点在于难以对研究课题提供满意的答案。探索性研究之缺乏确定性同抽样的代表性问题有关。这一问题将在第七章中加以详尽讨论。只有理解了抽样及其代表性问题，才能了解某一探究性研究是实际上回答了问题还是仅仅提出了解答问题的途径。

（2）描述性研究

许多社会科学研究的主要目的是对情况和事件进行描述。研究者把研究过程中观察到的情形描述下来。美国人口普查就是描述性社会科学研究的范例。人口普查的目标是对美国人口的各种特征作一正确的、准确的描述。竞选期间的盖洛普民意测验的目的在于对选民的投票倾向进行描述。商品市场调查的目的则是对购买或将要购买某项产品的消费倾向进行描述。社会科学中描述性研究的两个方面——描述的质量和概括性，将在第五章和第七章分别予以论述。

（3）解释性研究

社会科学研究的第三个目标是对事物作出解释。研究选举人的投票倾向是描述性研究；研究为什么有些人准备投甲候选人的票，另一些人准备投乙候选人的票则属于解释性研究。

虽然区别研究的三种目的很有必要，但应注意到多数研究是同时包含着这三种目的的。例如某研究者想对某种新的精神疗法做出评价。这一研究首先包括探索的性质，它勾划出这种治疗方法的效果，然后要描述治愈率，最后要解释为什么这种治

疗方法对某种类型的人比其他类型的人更为有效等等。

2 分析单位

在社会科学研究中，研究对象千差万别。在这里不是指研究题目的不同，而是指分析单位的差别。分析单位是研究中采用的基本单位，其最终目的在于集合它们的特征以描述由这些分析单位组成的较大的群体或解释一些抽象的事物。

（1）个 人

社会科学研究中最常用的分析单位是个人。通过对个人描述的聚合与处理过程来描述和解释社会群体及其互动。

概括性理解作为社会科学的规范认为，适用于一切人的科学发现才是最有价值的。但在实践中，社会科学家很少研究一切人。虽然有些比较研究是跨越国界的，但他们的研究一般局限于生活在一个国度的人们，甚至在更小的界定范围内进行。这些界定的群体的成员可作为研究的分析单位，他们包括学生、居民、选民、家长、教授等等。上述每个

词代表一种由个人组成的总体。以个人为单位的描述性研究一般旨在描述由那些个人组成的总体，解释性研究则为了发现存在于该总体中的社会动力。

（2）群体

社会群体也可作为社会科学研究的分析单位。以家庭作为分析单位的研究可以描述每个家庭的收入和是否拥有彩色电视，再把这些家庭累计起来，描述它们的平均收入和拥有彩电的比例。然后可以据此推断是否高收入家庭比低收入家庭拥有彩电的概率大些。

其他群体层次的分析单位包括：朋友圈子、街头帮伙、已婚夫妇、普查街区、城市、地域等。上述每个词也代表某个总体。街头帮伙意指由所有街头帮伙组成的总体。对街头帮伙的描述性研究可以是对其在某城市的地理分布的描述；对它们的解释性研究则可能发现大帮比小帮更容易卷入械斗。

当社会群体作为分析单位时，它们的特征可以从其成员的特征中抽取。如可以用家长的年龄、种族或教育程度来描述家庭；解释性研究则可以确定家长受过高等教育的家庭是否比家长未受过高等教育的家庭有更多或更少的子女。

（3）组　织

正式的社会组织也可以成为社会科学研究的分析单位。例如公司。对公司特征的分析可能包括雇员数量、年度净利润、财产总额、少数民族雇员比例等等。在描述性研究中，我们可以考查所有公司的平均净利润；在解释性研究中，可以确定大公司是否比小公司更多或更少地雇用少数民族雇员。

适合作为分析单位的正式社会组织还有教堂、大学、军队师团及超级市场等。

下列例子可以反映出研究单位的复杂性。如果研究那些雇员中学龄前子女多的公司是否比雇员中学龄前子女少的公司更有可能成立日托幼儿园，所用的分析单位是公司；如果研究学龄前子女多的雇员是否比学龄前子女少的雇员更倾向于选择有日托的公司，分析单位是雇员；如果研究家长的公司有日托的孩子是否比家长的公司无日托的孩子更愿意参加日托，分析单位则是孩子。

（4）社会产品

另一大组分析单位可统称为社会产品，或社会实体及其行为的产物。例如：书籍、诗歌、绘画、汽车、房子、歌曲、陶器、笑话及科学发现等。社

会科学家可以美术作品为分析单位，分析在俄国、中国和美国三个国家画家的作品中，哪国的作品表现出最强烈的工人阶级意识。也可研究某份地方报纸对地方大学的评论，对随时间推移报纸评论立场的变化作出描述或解释。评论即是分析单位。

另一类适用于进行社会科学研究的社会产品是由人际关系造成的。例如，可将婚姻区分为在人种、宗教上是混合的还是单一的，婚礼是宗教的还是世俗的，结果以离异告终还是持续的等等。同类事例还可举出法庭案件、交通事故、离婚、斗殴、劫机事件、种族骚乱等等。

（5）体系错误

体系错误是指用一个分析单位作研究却用另一个分析单位作结论。假定我们想研究某市选举中一位女候选人的支持者的成分。首先从投票记录中可查出各个选区投她票的人数，此外我们还有各个选区的普查数据。数据分析表明，人口平均年龄低的选区比人口平均年龄高的选区投这位女候选人的票要多。于是做出结论：年轻选民比老年选民更支持女候选人，即年龄影响对女候选人的态度。这样做结论时，我们已经大有可能犯了体系错误，因为在那些“年轻”的选区可能是老年选民投了女候选人

的票。问题出在我们用选区作为分析单位进行研究，却做出关于投票人的结论。

类似的例子还有很多。假如我们发现黑人多的城市比黑人少的城市犯罪率高，并不能因此作出黑人犯罪率高的结论；假如我们发现新教国家比天主教国家自杀率高，也并不能就做出新教教徒比天主教徒更容易自杀的结论。诚然，人们也犯从个人经验做出关于整体模式结论的错误，“体系错误”却是专指那种从对群体的观察作出关于个人的结论的错误。

（6）简约论

另一个与分析单位有关的问题是简约论。它是指在理解人类行为时采用过于有限的概念或变量。社会学家倾向于只考虑社会学变量（如价值、规范、角色）；经济学家只考虑经济学变量（如供需关系、边际效用）；心理学家只考虑心理学变量（如个性类型）。仅用经济因素解释全部或大部人类行为的作法称为经济简约论；仅用心理学因素解释人类行为的作法称为心理简约论，余可类推。

不同类型的简约论会导致在研究中使用不同的分析单位。社会学家把人们共同的价值观视为美国革命的原因，于是以个人为分析单位研究美国革

命；经济学家可能选择13个殖民地，以殖民地为单位研究它们各自的经济组织和革命条件；心理学家则可能选择单个领导人作为分析单位去研究他们的个性对事件进程的影响。

总而言之，由于很难确定对某项研究来说什么分析单位是最恰当的，社会科学家特别是不同领域的社会科学家常常为此争论不休。

3 研究题目

（1）状　态

关于研究题目的选择，首先可以研究不同分析单位的各种存在状态。对个人状态的描述可包括性别、年龄、身高、体重、婚姻状况、出生地等。对社会群体和组织状态的描述可包括规模、结构、及对其成员的概括性描述。对物质性社会产品的描述可包括大小、重量、颜色等，也可从它们与人发生关系的角度加以描述。对人际关系产物的研究则可包括对其在何时何地发生、涉及哪些人等状况的描述。

（2）意　向

以个人为分析单位的研究常以“意向”为研究对

象，如态度、信仰、个性特征、偏见、素质等。对个人意向的描述可包括信教的、政治上自由化的、反犹的、智力超群的、迷信的、科学的等等特性。对社会群体和正式组织的研究可包括其目标、政策、规范、过程及其成员的集体意向。对人际关系产物的研究内容也很丰富。例如可以把劫机事件区分为有政治性动机和无政治性动机的两类加以研究。

（3）行　为

社会行为亦可成为研究题目。我们可以直接观察某些社会行为，如投票、购买证券、投资、罢工、退学、上教堂作礼拜或买某种牌子的牙膏，也可对关于此类行为的第二手资料进行研究。对社会群体和组织行为的研究可包括全家外出郊游、祈祷、或搬家；公司对政治竞选的捐款、与其他公司合并、冻结价格或破产。人际关系的产物本身就是一种行为，似乎难以想象某种行为本身的行为。然而，婚姻这一行为可以有成功的，有失败的；法庭案件可以有证明有罪的，有无罪开释的；打架斗殴也可包括平息下去和失控等不同情形。

4 时间尺度

（1）横向研究

许多研究的设计集中在一个时间点上，研究某种现象的横断面。探索性和描述性研究多属于横向研究。例如，美国人口普查即是在给定的时间点上对美国人口状况的描述性研究。有些解释性研究也可以是横向研究。如关于种族和宗教偏见的全国性调查往往是在事物发展变化过程中取一个时间点来作的研究。

（2）纵向研究

另外一种研究称为纵向研究，即对一段时期内的情况进行研究。纵向研究包括三种特殊类型。

〔趋势研究〕对一般总体随时间推移而发生的变化的研究。例如，通过对不同时间点上的几次美国人口普查的比较看全国人口的增长趋势，或者从竞选过程中连续几次盖洛普民意测验结果研究各候选人的势头。

〔人口特征组研究〕对某一特殊人群随时间推移而发生的变化的研究。最典型的人群是年龄组，

如以三十年代出生的人为一个年龄组。除根据年龄分组之外，还可根据一般人口特征分组，如把越战期间大学在校生分为一组，把于1964年结婚的人分为一组等等。此类研究可举关于生于30年代初大萧条时期的人的经济态度的研究为例。这一研究的基本方式是每隔十年进行一次全国性调查。1950年，调查从20—25岁的人中抽取样本；1960年，从30—35岁的人中抽取样本；1970年，从40—45岁的人中抽取样本。虽然每个样本由不同的人组成，但完全可以代表生于1930—1935年的一代人。

〔定组研究〕定组研究与趋势研究和人口特征组研究相似，区别在于每次研究都用同一样本。以关于某次选举的研究为例，在整个竞选期间，研究者对同一个样本的人每月访问一次，问他们准备投谁的票。这一研究能精确地反映出人们意向变化的模式。

纵向研究明显优越于横向研究，因为它可以描述过程，但是这种优越性的代价是较多时间和金钱的耗费。

（3）近似纵向研究

在只有横向研究数据时也不是完全无法研究事物的过程。有时可以通过逻辑推理对过程作出近似

的结论。例如，关于学生吸毒的调查表明，有的学生吸过两种毒品，而只吸过一种毒品的学生全是只吸过毒品A而不是B，于是可以推论毒品A是先于毒品B出现的。此外，可利用横向研究中研究对象的不同年龄推论随时间推移而发生的变化。如研究人的生命周期中身体状况恶化的过程，可把研究对象的健康档案按年龄分组。从每个年龄组不同的健康状况可获得个人健康状况的近似发展史。近似的过程研究还可要求研究对象回忆一些事情，如出生地、高中毕业时间、1976年大选时投了谁的票等等。这种方法的危险在于记忆错误。在选举后的调查中，说自己投了当选者票的人数总是比实际投票人数多。

5 研究动机

（1）检验一般理论

有些社会科学家试图理解全部或大部社会现实。塞缪·斯托芬（Samuel Stouffer）曾试图检验帕森斯（Paosors）关于行为的一般理论。帕森斯的理论包括评判社会行为的标准。他认为在某些情况下应当运用普遍性的标准，如法官判案时；在

另外一些情况下应当运用特殊性的标准，如家长可用不同标准对待大孩子和小孩子。他指出，普遍性与特殊性标准可能发生冲突，如警察发现自己的儿子抢劫或违反交通规则时会陷入两难境地，作为警察的角色和作为父亲的角色会发生冲突。

斯托芬决定研究角色冲突的问题。他的调查是在哈佛大学的学生中进行的。他提出的问题是：假如你发现你的朋友犯了法，你是根据对待罪犯的普遍性原则去做呢，还是按朋友义气的特殊性标准去做？他的研究创造了用具体标准检验抽象理论的范例。这个事例还体现了社会科学研究中的劳动分工：一些社会科学家创造理论，另一些人搞实验研究，以检验和重新定义他人的理论。这种分工促进了社会科学的发展。

（2）检验具体假设

除了以检验综合性理论为目的的实验研究之外，还有大量注重发展和检验具体理论假设的研究。如，研究社会上哪一种人参加爱国主义运动的积极性最高。检验综合性理论的方法也适用于具体假设的检验，区别在于综合性理论检验的结论要归结到理论上，而具体假设检验的结论归结到假设就可以了。

（3）无框架研究

科学家常常会选中一个有趣的题目但对变量之间的关系并无清晰的看法。许多人认为这种无框架的探索性研究不如事先作好框架的研究，其实无框架研究也许会比有框架研究得到更丰富的成果。在事先作好框架的研究中，分析只能肯定或否定预期的假设；而在无框架研究中往往会得到始料不及的发现。有框架研究的缺点在于容易忽略那些未预料到的关系。这并不是说社会研究者可以无计划地乱搞，而是为了使人们注意到科学研究有多种途径，不一定总要有事先作好的假设。

（4）合同研究

社会科学家与某机构签订合同从事某一特殊问题研究的现象变得日益普遍。这类研究多属实用性质。地方政府可能委托研究者搞关于失业率的调查；公司可能委托研究者对其训练项目作出评价；竞选人可能委托研究者去搞民意测验。虽然这类合同研究偶尔可能对科学的基本知识有所增益，但多数情况下它只是为具体行动服务的研究。

6 研究设计

图4—1提供了社会科学研究过程的图解。图的最上部是兴趣、看法和理论，研究的可能的起点。字母A、B、X、Y等代表变量或概念，如偏见：研究可能对某些人比其他人偏见程度高这一现象发生兴趣。研究也可从对某事的独特看法或现存理论开始。

〔概念之定义〕当我们想到一个概念（如偏见）时，其内涵外延往往并不清楚。必须首先明确地定义所有的概念，然后才能开始进行研究。

〔研究方法之选择〕社会科学研究有多种方法，每种方法都各有利弊。对某些概念的研究用某种方法比用其他方法更合适。根据对《美国社会学评论》杂志文章的研究（见表4—1），从1936到1978年间，早期研究采用解释方法者（其中包括对文献、书籍的历史研究、内容分析、观察、无框架访问等）与采用调查方法者各半。随时间推移，用调查方法的比例越来越大（从48%到80%），用解释方法的比例越来越小（从51%到17%）。用实验方法的始终很少（2%）。

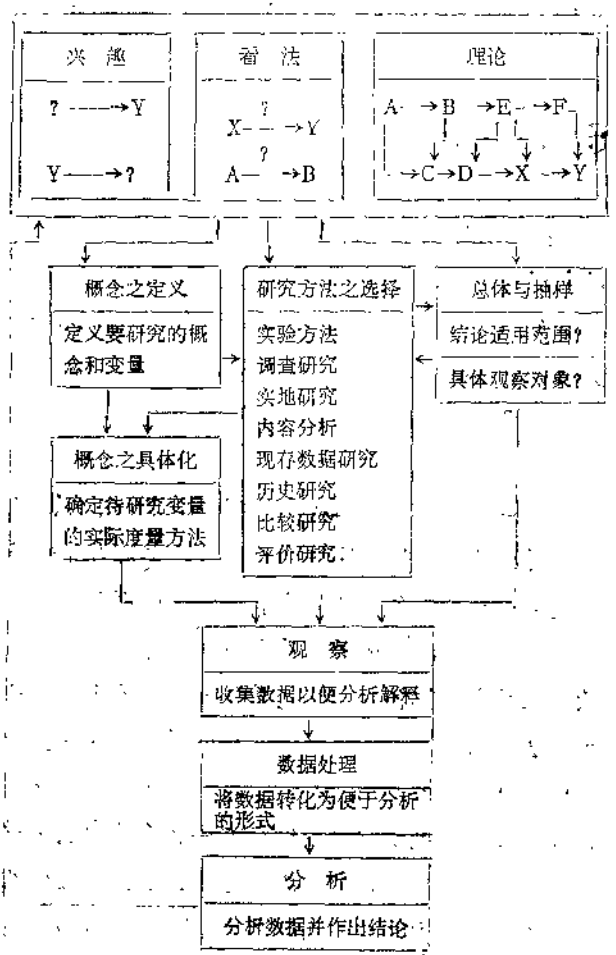

兴趣
?——→Y
Y——→?
看法
?
X- - →Y
?
A— →B
理论
A→B→E→F
↓ ↓ ↓ ↓
→C→D→X→Y
概念之定义
定义要研究的概念和变量
概念之具体化
确定待研究变量的实际度量方法
研究方法之选择
实验方法
调查研究
实地研究
内容分析
现存数据研究
历史研究
比较研究
评价研究
总体与抽样
结论适用范围?
具体观察对象?
观察
收集数据以便分析解释
数据处理
将数据转化为便于分析的形式
分析
分析数据并作出结论

图4—1 研究过程

表4—1 《美国社会学评论》*杂志文章分类

	年代			总计
收集数据方法	1936—1949	1950—1964	1965—1978	1936—1978
解释方法	51%	27%	17%	31%
调查方法	48%	70%	80%	67%
实验方法	1%	3%	3%	2%
	(137)	(200)	(152)=(489)=100%	

* 译注：《美国社会学评论》杂志是美国主要社会学杂志之一。

〔概念之具体化〕在定义了待研究的概念并选择了适当的研究方法之后，我们必须制定具体的度量尺度。概念之具体化意为制定度量特殊概念使研究得以实施的具体步骤和作法。

〔总体与抽样〕除了定义概念和设计尺度，还要确定具体研究对象。研究的总体是指研究将从中抽取结论的群体。我们常常不能直接研究总体中的一切成员，而只能进行抽样研究。第七章将讨论使样本能精确反映总体的抽样方法。

〔观察〕在确定了内容、方法和对象之后，即可进行观察和收集数据了。本书第三篇将介绍各种不同的观察方法。

〔数据处理〕第十三章将描述社会科学数据处理的方法。

〔分析〕研究过程的最后阶段是从数据中得出

结论。第十四章到第十八章提供了多种数据分析的方法。数据分析的结果将反馈到研究之初的兴趣、看法和理论，这一反馈将成为新的研究过程的起点。

综上所述，研究设计包括确定研究的目的、方法、及研究内容。研究设计的基本考虑有三点：兴趣、能力、以及能够利用的资源。

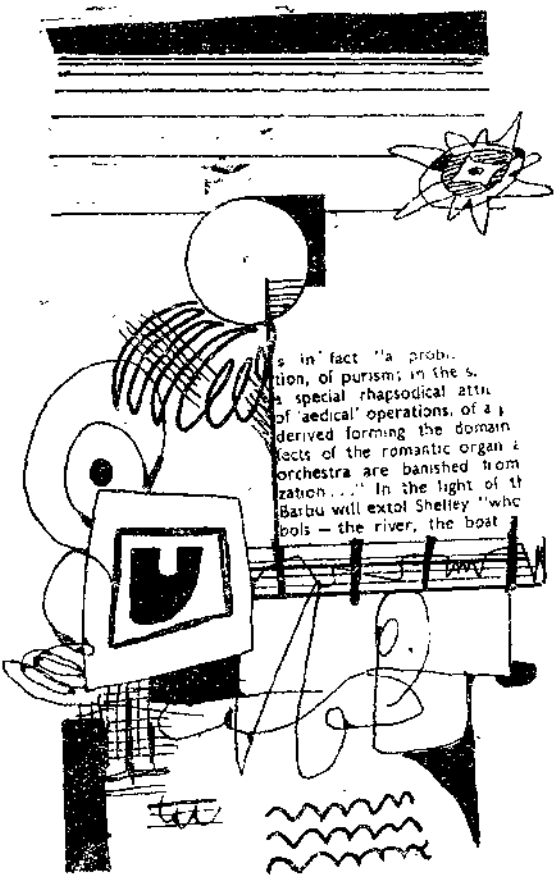
一般研究设计过程是这样开始的：研究者首先自问对什么问题最感兴趣。人们对种种社会现象肯定会有不解之处：为什么有些人在政治上持自由派观点而有些人持保守派观点？为什么某些人比其他人宗教信仰强烈？大专院校是否仍然存在歧视女教师的现象？种族混合的婚姻是否比其他形式的婚姻更巩固？学生的学习效果是大班上课较好还是小班上课较好？美国经济比过去是更多还是较少地依赖于战争与国防需求？应当认真考虑哪类问题对自己吸引力最大。

兴趣确定之后，则应考虑解答问题所需的信息及能够提供适当信息的分析单位：大学在校生、公司、选民、城市还是其他？如何得到这些信息？有关这一课题的数据或许已经存在（如在政府文献中），如果需要自己收集资料，应当采用哪种方法去收集？有无必要抽取样本？是否可以通过参加一

些组织的会议了解情况？是否能从图书馆收集到研究所需的数据？

回答这些问题的过程就是研究设计的过程。要衡量自己的研究能力和财政情况，不要试图设计一个完美但无力实行的研究计划。

一旦确定了研究内容及方法，即应查阅书刊杂志中前人已做过的研究，看关于这个题目他人已做了些什么。文献考察可能导致研究设计的修正；可以采用某种他人使用过的方法或索性重复他人做过的研究。对某个研究课题的独立重复是自然科学中最常见的作法，这一点对社会科学来说也很重要，但人们常常忽略这一点。此外，可以研究这一课题中被他人忽略的方面，或采用与前人不同的方法。用不同方法检验同一发现称为“三角测量”，是社会科学研究中一个很有价值的研究策略。



第五章
概念之定义与度量

1 现存事物之度量

当人们读过一些关于宗教信仰程度或种族偏见程度的研究论文之后，往往对研究者所使用的度量方法发生怀疑，认为它们过于表面化，没有表达出被研究的概念的本意。例如研究者只用上教堂的次数来度量宗教信仰程度；用人们在某一次选举中投了谁的票来度量其政治倾向。前面曾经说过，观察是科学的两大支柱之一。由于观察一词也可指一种偶然的被动的行为，科学家用度量一词取而代之，意为对现实世界的认真而有意识的观察，目的在于运用构成变量的事物的不同性质对事件和对象作出描述。虽然度量在社会科学中有其特殊的问题，但我们能够度量一切现存的事物。换言之，任何事物只要存在即可度量，无一例外。

（1）什么是偏见？

什么是偏见？偏见这个东西并不存在。我们既不知道它的形状、体积和颜色，也从未摸到过它，但是我们确实常常谈起它。我们注意到对黑人施加私刑的人称黑人为黑鬼，他们之中的许多人还歧视妇女。我们感觉到世界上有一些人具有上述几种倾向。当谈到这类人时，我们说："这个人就属于那种对黑人施私刑、称黑人为黑鬼、不愿意雇用妇女当经理的人。"听来多么啰嗦。于是一个好主意油然而生："我们用偏见这个词来形容这种人吧。"这就是偏见一词的来历。"偏见"本身并不真正存在，我们只是用它作为谈论某种人的捷径。归根结底，偏见只是我们相互约定的用于思想交流的一个用语，我们用它来代表在现实生活中所观察到的许多明显相联的现象的集合。

（2）概念之形成

虽然字典固定了社会对每一用语的约定用法，每个人在这些约定用法上仍会加上自己的想象。由由每个人有不同的经历和观察，没有任何两个人对某一概念的理解是完全一致的。

如前所述，任何实存的东西都是可以度量的。

例如，我们可以度量派特是否把一只掉出鸟窝的小鸟放了回去，他是否在圣诞节去医院看望过病孩，他是否曾被电影感动得流泪等等。这些事都是确实存在的，因此我们可以度量它们。但对派特是否富于同情心这一问题我们却不能回答，因为不能象度量现实中确实存在的事物那样度量“同情心”，它同前面提到的那些事情有本质的区别。

同情心作为一个词是确实存在的。它由三个字组成，以“同”字打头。这是同情心这个词确实存在的一面。与同情心这一概念有关的某些其他方面也是确实存在的。例如，当一个小学教师问班上的小学生有谁知道“同情心”一词的含义时，有几个人举手这件事就是确实存在的。我们不能度量“同情心”的含义，因为它不是确实存在的事物，它只存在于一种约定的形式之中，这一约定是关于在思想交流中如何运用一个词描述确实存在的事物的约定。

在这个意义上，阿布拉汉·卡普兰（Abraham Kaplan）将科学家度量的事物分成三个层次。第一层次是可以直接观察的事物，如苹果的颜色或问卷上的记号。第二层次是可以间接观察的事物，如通过一个人在问卷上圈了“女”字可以间接地观察到这个人的性别。最后一个层次是思维的产物，即那些建立在观察基础上但是不能被直接或间接观察

到的理论性创造的产物。例如平均分是学生各门功课所得分数的数学概括。卡普兰将概念一词定义为思维的产物。“同情”这一概念就是所有使用这个词的人们的思维的产物。它不能被直接或间接地观察到，因为它并不存在，是我们创造了它。

（3）概念之定义

人们是通过对词语用法的一个含糊而笼统的约定系统来进行日常交流的。我们常常并不确切理解所用的词语，但能大概明白所要表达的意思。虽然人们对于“同情心”一词的理解不完全吻合，但至少可以肯定“富于同情心”的派特不会去撕掉昆虫的翅膀。大量的误解和冲突——从人与人之间到国与国之间——是我们为这种不精确所付出的代价，但我们还能混得过去。然而科学却容不得半点含糊和不精确。

概念之定义意为对在什么意义上运用某一特殊用语作出精确的说明。假定我们想了解女人是否比男人更富于同情心，如果不确定在什么意义上运用“同情心”这一用语，研究就毫无意义。

（4）指标与尺度

定义概念的最终产物是表达我们意思的一组指

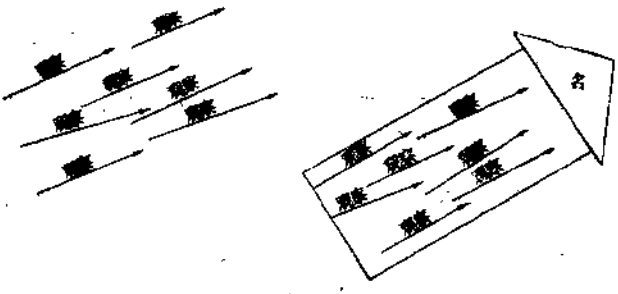
标。这些指标是确实存在的可以观察到的事物。如，我们可以约定用圣诞节去医院看望病孩作为有同情心的一个指标。用把掉出鸟窝的小鸟放回去作为另一个指标等。我们设十个类似的指标，并通过统计发现派特有六条，约翰有三条，玛丽有九条，且发现女性的平均数为6.5，男性为3.2，从而做出结论：女人比男人更富于同情心。

我们还可以运用不同的尺度把同情心分为感觉上的同情和行动上的同情，宗教上的同情和政治上的同情。这种分类使我们对事物的理解更加深刻。例如我们可能发现女人比男人在感觉上更富于同情心，而男人比女人在行动上更富于同情心，或者恰恰相反。

（5）定义与现实的混淆

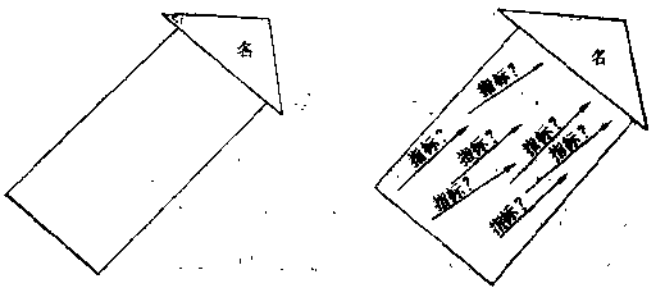
让我们作一简单回顾。概念是由想象而来的，想象是由看去有联系的观察和经验综合而成的。观察和经验是确实存在的，概念却只是思维的产物。表达概念的用语仅仅是为交流思想而设计出来的。归根结底，词语只是字的集合，并不意味着实存。

然而，人们常常由于把概念当作具有实存意义的事物而陷入陷阱，尤其当我们认真地对待某个概念并精确地使用它时，就更加危险。一旦假定概念



1、在现实生活中，许多观察似乎有共同之处，我们意识到在个别之上有一般，并进一步发现可以用概念来交流思想。

2、每当谈到那些具有共同概念的个别现象时却对它们作一番描述是很不方便的，于是我们为这个概念命名。



3、当我们用这个概念交流思想时，渐渐以为它是确实存在的而不仅仅是现实世界中某些具体现象的总结了。

4、由相信概念本身确实存在到反过来讨论哪些现象才是某一概念的恰当指标，这样的结果是我们始料不及的。

图5—1　名实渐清的过程

有实存的意义，我们又会为发现其实存的意义和恰当的度量尺度而大受周折。图5—1表达了这一过程的历史。我们对观察到的确实存在的事物做出概念的概括是为了用起来方便。然而由于常常使用它们竟致于认为它们是确实存在的了。这个弄假成真的过程又称为名实混淆，而概念的名实混淆在日常生活中是司空见惯的。

2 定义与研究目的

在第四章我们曾谈到过两种研究目的：描述和解释。这两种目的之间的区别造成了定义和度量方法的不同。一般认为描述性研究比解释性研究简单，但是在描述性研究中定义要困难得多。请看一个描述性语句："45%的学生在政治上是保守的。"试图在这个句子用语的定义上取得一致是相当困难的，而比例数字的大小直接依赖于对所度量事物的定义。对"政治上保守"的不同定义完全可能导致另一个结论："5%的学生在政治上是保守的。"

定义在解释性研究中反而比较容易。假定我们想解释为什么有些人比其他人在政治上更倾向于保守主义，进一步说，假定我们想了解老年人是否比

年轻人一般来说更保守些。我们可以准备25个关于“保守”的具体指标，尽管我们对哪些是最佳指标尚存分歧，但当我们发现在25个指标上老年人的回答全都比年轻人保守程度高一些时，究竟哪些是最佳指标的问题已经不重要了。虽然我们在“保守主义”的定义上始终有分歧，仍可做出老年人一般来说比年轻人保守这一结论。结论的这种一致性在描述性研究中就很难成立，定义的改变毫无疑问会导致不同的结论。

3 度量的质量

社会科学家能够度量任何确实存在的事物，但许多待研究和度量的事物实际上并不存在。度量这些不存在的事物也是有可能的。本节将讨论如何评价度量的质量，其中包括对那些并不存在的事物的度量。

首先，度量具有不同的精确度。描述某人为43岁比说四十多岁更精确；描述某物长95厘米比说长一米左右更精确。一般来说，精确的度量优于不精确的度量，然而精确并不一定总是必要的。如果在一项研究中了解一位女性是四十多岁就够了，任何

更精确的调查就是浪费精力。因此概念的具体化应当对精确度作出规定。如果不能确定精确到什么程度合适，则应精确一些。

不应混淆精确和正确。描述某人“生于纽约”比说他“生于美国”更精确，但如果他实际上生于华盛顿，比较不精确的描述反而更正确了。显而易见，精确和正确是衡量度量成败的重要标准。另外两个重要的标准是可靠性和准确性。

（1）可靠性

可靠性是指将某种特殊技巧重复应用于同一题目的研究，每次都能得到同样的结果。假定让甲猜乙的重量，甲开始猜165磅，当乙化了装后又让甲猜，甲猜180磅。这说明甲对重量的估计不可靠。然而当甲得到一个磅秤以后，无论乙化装多么巧妙，称量的结果都是一样。这说明磅秤比起猜测是度量重量的更好工具。

可靠并不意味着正确。假如把磅秤调低5磅，虽然每次都能称出同样的重量，但次次都是错，这一差数称为偏差。

许多社会研究方法中都存在可靠性的问题。在调查研究中，由于访问者的态度和所关心领域的不同，会得到不同的访问结果。有时研究对象根本不

知道问题的答案（如问“你一共去过多少次教堂？”）；有时问卷提出的是研究对象不关心的问题（如“你对芬兰和土耳其目前的关系是否满意？”）；有时则由于问题太过复杂，人们会对问题做出不可靠的回答。

如何保证度量的可靠性？首先，当提出问题时要注意只问那些被访问者可能知道答案的问题。问题要提得清楚明白。最大的危险在于人们会给出一个回答，如对芬兰土耳其关系的看法，而实际上对这两国的关系一无所知。

其次，研究者可用同样或略有差异的提法将一个问题反复多问几次。如，为度量异化程度可提出十个问题，如果人们对其中一个问题的回答与对其他九个问题的回答大相径庭，就应将这一问题删去。

此外，可在不同的时间点上提出同一问题，应当特别注意的是，被访问者自身的改变可能引起研究者对自己所收集数据的可靠性的怀疑。例如有人在一项研究的过程中结婚了，使研究者因此对婚姻状态数据的可靠性产生了怀疑。

（2）准确性

准确性这一术语通常是指实验的度量标准反映某一概念的真正含义的程度。

首先，关于度量的“表面的准确性”。某些特殊的实验度量指标可能与人们对某一概念的个人想象相符或不太相符。人们可能在某一指标是不是某一概念的最佳指标这一点上发生争论，如用工人到工会发牢骚的次数当作衡量工人劳动积极性的指标，但至少会认为发牢骚次数与劳动积极性有关。假如有人提出用工人业余时间借阅书籍的数量作为度量劳动积极性的指标则无疑会招致反对：这一度量指标没有任何表面的准确性。

其次，对许多概念的度量研究者已取得一致意见。例如普查机构已经制定了大量概念的具体度量指标，如家庭、就业状态等概念的度量指标。

此外，某项指标与另一指标在同一概念上产生同样结果可以说明这一指标的准确性。本书第四篇还将专门讨论这一问题。

图5—2用图示表明可靠性与准确性的区别。用打靶来比喻对事物的度量，可靠性象一组密集弹着

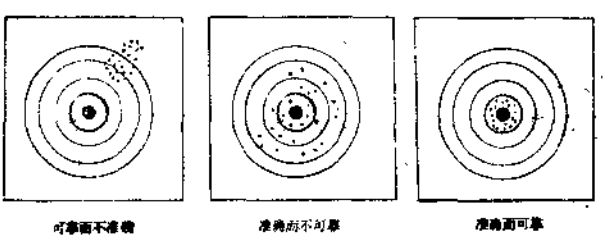


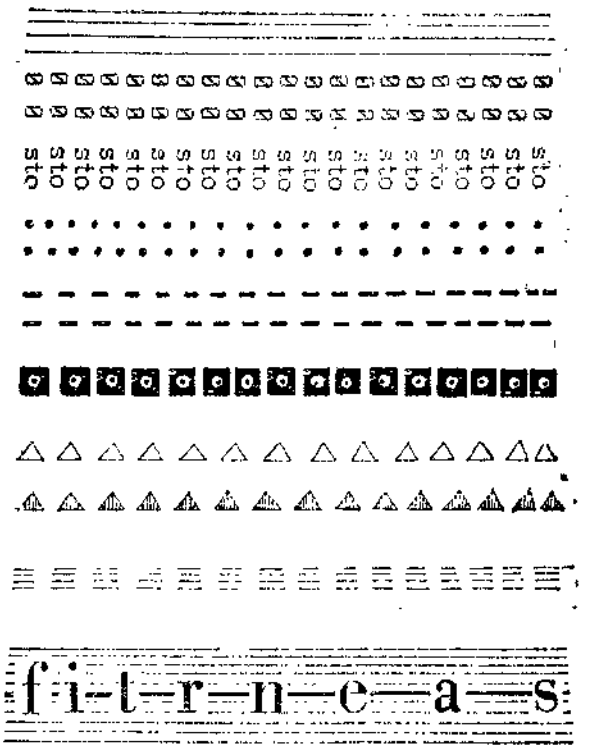
图5—2　关于可靠性与准确性的一个比喻

点，无论打在靶子的哪个部位；而准确性则要求打在靶心附近。从图中可以看出，可靠而不准确属于系统误差；准确而不可靠则属于随机误差。

（3）可靠性与准确性之间的相对关系

可靠性与准确性之间存在着某种程度的相对关系。例如，不去统计工人到工会发牢骚的次数，而在生产线上作直接观察，或与工人交谈，关于工人劳动积极性的研究会更具准确性。然而用统计工人到工会发牢骚次数的办法则可以使关于劳动积极性的研究更具可靠性。

科学要求可靠的度量，但可靠的度量却往往破坏了某一概念含义的丰富性。如劳动积极性的高低是远远不能由去工会发牢骚的次数多少这一点来表达的。这是科学研究中不可避免的两难处境，应当对这一点作充分的思想准备。假如人们尚未对如何度量某一概念取得一致意见，就采用几种不同的方法来度量它。假如某一概念可以用几种不同的尺度来度量就把每种尺度都试用一下。总而言之，概念在我们给它指定的意义之外没有任何其他意义，而对人们为某一概念所赋予的特殊意义作出评价的最终标准值在于其有用性，即能否帮助我们更好地理解周围的世界。



第六章
概念之具体化

1 概念之具体化

（1）差异范围之界定

概念具体化过程的基本点是对差异范围之界定，即为变量的取值分组。例如，我们要研究人们的收入。岁入最高的人可能达到数百万，但多数人没有那么多。除非打算研究巨富，没有必要把高收入组的起点定得太高。我们可以把高收入组的起点定为五万美元，亿万富翁和岁入刚好五万的人被分在一组对我们的研究并无影响。在美国低收入组一般界定为年薪五千美元以下。

在对态度和倾向的研究中，差异范围界定的问题有其特殊性。如果粗心大意可能导致遗漏大量信息。例如，我们想了解人们对发展核动力发电的态

度。我们推测对这一发展项目有人喜欢有人不喜欢，于是把答案的差异范围界定为从“非常喜欢”到“非常不喜欢”。问题出在有一部分人不是不喜欢而是根本反对发展核动力发电。正确的作法应当把范围界定为从非常喜欢到无所谓到强烈反对。

在社会科学研究中，对许多变量都应当做类似考虑，例如对于宗教，人们不仅有信仰强烈薄弱之分，还有根本反对宗教信仰的。这并不意味着必须在每个研究中都运用差异的最大范围，差异范围的界定应依研究目的而定。如果搞清不信宗教和反对宗教的区别对某项研究无用则不必去度量它。此外，度量的范围还依研究对象而转移。如研究人们对高等教育的基本看法，当研究对象是大学教授时，或许不必考虑“高等教育危害学生健康”这种观点；而如果研究对象是学生则另当别论。

（2）精确度问题

在考虑变量的具体化时也有一个精确度的问题。某项研究是必须搞清某人为17岁、18岁，还是把他分在10—19岁这一组就可以了？如果研究题目是有选举权的人参加投票的比例，研究对象究竟是17岁还是18岁就显得很重要。因此精确度应当依研究目的而定。当研究人们参加政党的情况时，是必须

搞清某人是保守民主党人、激进民主党人，还是了解到他属于哪个党就可以了？关于婚姻状态的研究，是仅仅了解某人结婚与否就可以了，还是要详细区分，从未结婚、分居、丧偶和离婚几种情况？对这些问题并无一般答案，答案依研究目的不同而不同。一个有用的原则是当不确知度量应当详细到什么程度时，详细比简单好。在数据分析过程中，我们可以把细致的分类合并成粗略的分类却不能把观察和度量时混在一起的内容再作细致分类了。

（3）研究角度问题

当人们制定变量的具体尺度时常常并不确知自己究竟想了解变量的哪一方面。假定我们研究人们对政府的态度及对政治腐败的看法，可能的研究角度有：政府是否有腐败现象，腐败到什么程度，导致腐败的原因，腐败现象是否不可避免，应当如何对待腐败现象，为防止政府腐败人们打算采取什么行动，等等。人们对政府腐败的认识是多方面的，因此明确研究角度是非常重要的。

（4）度量层次问题

概念之定义及具体化过程也是确定变量及其具体属性的过程。如就业状态即是由就业与失业两个

属性组成的变量。

变量具有两个重要的性质。第一，构成变量的取值应当是穷尽的。每个被观察到的现象都应属于某一类取值。如果党派这一变量只设共和党和民主党两个取值，那些参加了社会主义工人党、美国独立党、和平与自由党的人和无党派人士将无类可属。如果在共和党和民主党之外加上“其他”和“无党派”，变量取值就是穷尽的了。无论如何设计，必须使每个观察现象都能被归入某一类之中。

第二，构成变量的取值必须是相互不包容的。每个观察现象都应属于一个取值且只属于一个取值。例如当定义就业和失业时，不能使任何人同时具有这两种属性。应当把有工作同时在找另一工作的人归入哪一类呢？在这种情况下，可以按就业优先的原则即只要这个人有工作就算具有就业这一属性无论他是否在找另外的工作。

变量有不同的量度层次，其中包括定名的、定序的、定距的和定比的。

〔定名变量〕那些取值只包括穷尽和相互不包容性质的变量属于定名变量。例如性别、宗教、出生地、大学专业和头发颜色等。我们将一群人按出生地分组：生于纽约的为一组，生于华盛顿的为另一组，等等。所有分在一组的人具有至少一个共同

点，其他组的人在这一点上与之相区别。每组之间的距离远近是无关紧要的。

〔定序变量〕那些取值可以按某种逻辑排列顺序的变量属于定序变量。例如社会阶级、信仰程度、保守主义、异化程度、偏见程度、智力程度等等。将上述同一群人按其教育程度分组：大学毕业生为第一组，中学毕业生为第二组，未从中学毕业者为第三组。这种分组满足穷尽与相互不包容的条件，加上可以按照教育程度高低对这三组作逻辑排列。如果我们知道两个人分别在哪一组，即可确定一个人比另一个人受过较多、较少或相同的教育。

对定序变量来说，组间的距离是无关紧要的。第一组距第二组可能是五米远，第二组距第三组是五百米远。距离在这里毫无意义。但中学组必须在大学组和低于中学组之间，否则顺序就错了。

〔定距变量〕那些取值间的距离具有实际意义的变量属于定距变量。如华氏或摄氏温度刻度。80度和90度之间的区别与40度和50度之间的区别相同。但是我们不能说80度比40度的天气热两倍，因为温度表上的零度是人为设置的。零度并不意味着没有温度。智商100与110的间距同110到120的间距相同，但智商为零的人并不是没有智力，虽然他可能不适于做大学教授。

〔定比变量〕定比变量除具有上述三种变量的全部性质之外，还有一个具有实际意义的零点。例如年龄、在某地居住的时间、加入了几个组织、某段时间内上教堂的次数、结婚的次数等等。将研究对象按年龄分组：一岁的一组、两岁的一组、三岁的一组等等。每组成员岁数相同且与其他组不同，这一点符合定名变量的条件；这些组可按顺序排列，这一点符合定序变量的条件；相邻两组距离相等，这一点符合定距变量的条件。最后，由于具有实际意义的零点（在母腹中尚未出生的婴儿）成为一个取值，又符合定比变量的条件。因此可以作出某人年龄比另一人大一倍的描述。（参照图6—1）

〔度量层次的运用〕特定的分析技巧适用于特定的度量层次。例如，一群人的平均年龄是可以计算的，但其平均宗教信仰是不可计算的，因为宗教信仰是定名变量。然而，有些变量也可按不同变量层次来处理。定比变量层次最高，依次下推为定距、定序和定名。高层次的变量可以作为低层次变量来处理。例如可将年龄作为定序变量来处理，分为青年、中年和老年；也可将其作为定名变量来处理，以是否生于大萧条的三十年代为界，将研究对象分为两组。

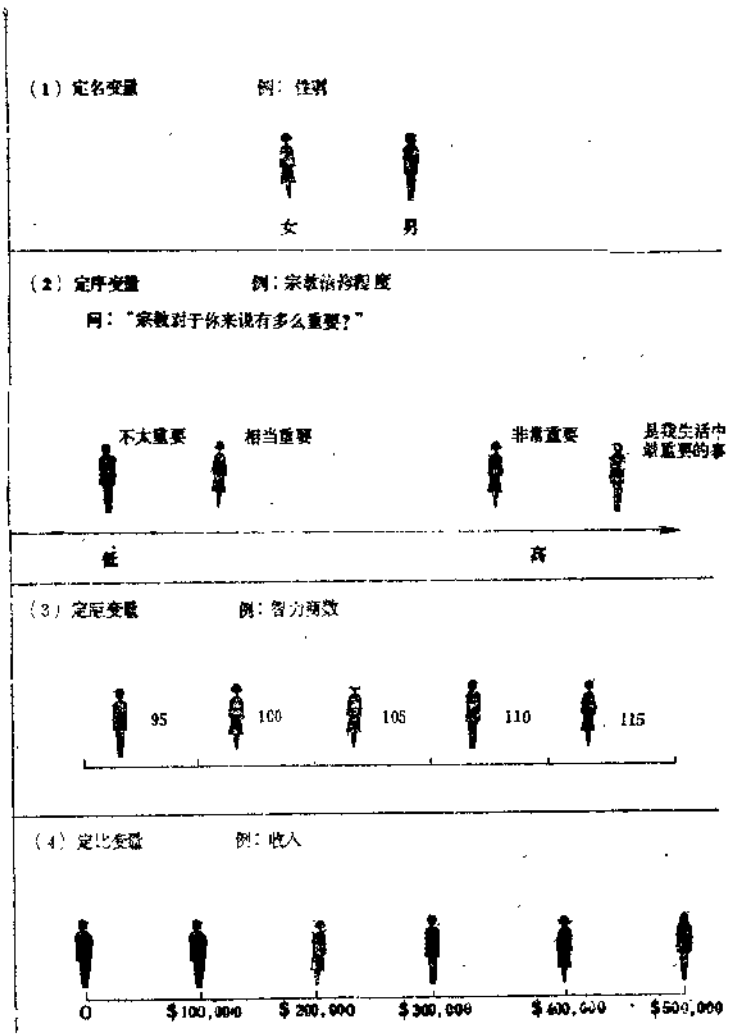

（1）定名变量
例：性别
女
男
（2）定序变量
例：宗教信仰程度
问："宗教对于你来说有多么重要？"
不太重要
相当重要
非常重要
是我生活中最重要的事
低
高
（3）定距变量
例：智力商数
95
100
105
110
115
（4）定比变量
例：收入
0
$100,000
$200,000
$300,000
$400,000
$500,000

图6--1 度量层次

（5）单指标与多指标

许多社会科学的变量具有相当简单明显的度量标准。无论怎样分，性别总是男女两种属性。此外如家庭拥有孩子的数目。大量变量具有这样明显的单指标。

然而也有一些变量不能用单一的指标来度量。许多概念有多种解释，每种解释又有多种可能的度量指标。在这种情况下，对一个变量要做多种观察并制定综合的度量标准。例如对学习成绩这个变量来说，每门课的分数都是一个潜在的指标。只用一门课的分数不可靠，于是我们用各门课的平均分来代表学生的学习成绩。社会研究中常需要制定这类综合度量指标。

2 概念具体化范例

这里介绍几项特殊的研究课题及可能采用的研究方法：

（1）女人是否比男人更富于同情心？

a. 选择一组研究对象，对他们描述人们遇到

危难的一些假想的情况，问他们如果碰到这种事会如何处理。例如，如果他们碰到一个迷路的孩子哭着找妈妈，他们将如何处理。从对这一问题的回答来比较男人还是女人更富于同情心。

b. 搞一个实验。让一个孩子假装迷了路，观察是男人还是女人更多地提供帮助，同时要记清路经此地的确切人数，防止男女比例基数不明。

c. 选择一个样本，调查研究对象所加入的组织。统计男人还是女人更多地参加具有慈善性质的组织。可先分别计算慈善组织在每人所参加组织中所占的比例，再看男人还是女人的平均比例数高。

d. 从地方报纸上找一条能引起同情心的消息，如杀害小海豹，看随后几天就此消息寄给编辑部的读者来信中是男人还是女人更多地对此表示同情。

（2）社会学学生还是心理学学生对国际时事更加了解？

a. 对情况相似的两个班——一班是社会学学生和另一班是心理学学生——做国际时事小测验。

b. 向国际时事课教师索取这两个班学生的平均分数。

c. 要求这两个班学生分别在“呼吁将联合国总部迁往纽约”的请愿书上签名，计算一下每个班

有多少学生签了名，多少学生知道联合国总部就设在纽约。

（3）纽约还是加利福尼亚更适于居住？

a．查阅有关统计资料，比较两州的移进移出人口率。

b．查阅盖洛普等民意测验机构的调查结果，它们经常询问人们认为在哪个州生活最好。

c．比较两州的自杀率。

上述举例旨在开阔视野，使读者认识到可以通过多种途径、多种方法来度量变量。

3 提问规则

无论是调查、实验和实地研究，还是其他形式的研究都需要向研究对象提出问题，有时是由访问者提问，有时是由研究对象自填问卷。有几项规则可以帮助研究者正确地提出问题，避免得到无用甚至错误的信息。

（1）提问与陈述

问卷是一组问题的集合。问题的基本形式有两

种：一种是直接提问，另一种是陈述某一观点，然后运用莱克特（Likert）尺度度量研究对象对这一陈述的态度（强烈赞成、赞成、不赞成或强烈不赞成等）。同时运用提问和陈述两种方法可使问卷中的问题设计显得灵活有趣。

（2）开放型问题与封闭型问题

提问方法亦有两种：开放型问题要求答问者自由提供答案，如问“你认为美国当前面临的最大问题是什么？”封闭型问题是由答问者在研究者提供的答案中选择一种。由于封闭型问题便于进行数据处理，许多研究者喜欢采用封闭型问题这一形式。开放型问题在进入数据分析之前必须重新分类整理，其中有些答案可能全无用处；封闭型问题则可直接进入数据处理过程。

封闭型问题的主要弱点产生于研究者的答案结构。有时研究者提供的选择答案会有重大遗漏。如关于“美国面临的最大问题”，某些研究对象心目中最重要的问题可能并未被开列在选择答案之中。制定封闭型问题要遵守前面提到过的两项原则：第一，供选择的种类必须是穷尽的：囊括一切可能的回答。为了保证这一点，研究者常常在多种选择后加上“其他（请具体说明）”一栏。第二，供选择

的种类必须是相互不包容的：使研究对象只能选择一种答案。有时可以要求答问者在多个可能入选的答案中选择一个最佳答案。

（3）问题要清楚明确

问题必须清楚明确。有时由于提问者本人对问题很熟悉因而忽略了答问者对问题的熟悉程度。如“你对反弹道导弹系统的提案有何看法”这一问题可能引起反问：“哪个反弹道导弹系统的提案？”问卷中所提出的问题应当十分精确，使答问者确切了解研究者想问的是什么。

（4）避免双重提问

研究者有时会提出包含双重内容的问题。例如要求人们对以下说法发表意见：“美国应当放弃空间计划把钱用在国内发展上。”虽然有些人明确地表示赞成或反对，有些人却会认为无法作答。其中一部分主张放弃空间计划但把这些钱还给纳税者；另一部分主张继续发展空间计划但国内发展也应适当增加投资。对这两种人来说，无论填同意还是不同意都不符合本意。

（5）使答问者有能力作答

提出问题时要考虑答案的可靠性。某研究者坚持要一群司机回答自从拿到行车执照后到底行车多少里，尽管司机根本记不清楚他仍坚持要人作答，结果得到的回答是行驶了“成千上万里”。

（6）问题要简短

为了不至引起误解，研究者往往把问题搞得很长。应当避免长问句。答问者往往不愿意花时间研究问题的含义，应当使他们能很快看完、看懂、毫无困难地选择答案，因此问题应当尽量清晰简短。

（7）避免用否定形式提问

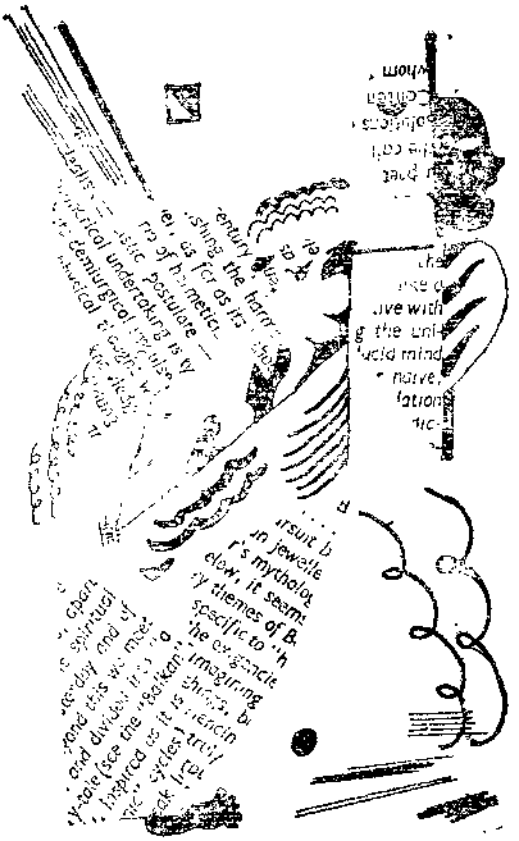
用否定形式提问题容易引起误解。当人们被问到是否赞成“美国不应当承认古巴”这一观点时，很多人会忽略那个“不”字。因此，许多赞成承认古巴的人填了“赞成”，不赞成承认古巴的填了“不赞成”，结果是永远也搞不清谁是什么意见。

（8）避免偏颇的问题

人们对问题的回答很大程度上依赖于问句的措辞。有些问题的提法似乎在鼓励人们做出某一种回

答，这样的问题称为偏颇的问题。例如在问句中加上“你是否同意美国总统提出的……”、“你是否同意最高法院决定的……”等字样，就容易得到肯定的答复。而用“你是否同意希特勒所主张的……”就容易得到否定的答复。

研究者要注意所谓“社会期望”的问题。无论提出什么问题，人们的答案总会经过“社会期望”的过滤，尤其在面对面调查的情况下。例如，某人实际上主张女人应当回到厨房去，不应当有选举权等等，但当被问到是否赞成男女平等时，他因为不愿意被视为大男子主义者而回答“是”。要避免提出使答问人感到为难的问题以及使其答案听上去不正当、不近人情、愚蠢或不负责任的问题。



第七章
抽 样 逻 辑

1 抽样的历史

社会科学中的抽样方法是与政治选举的预测一起发展起来的。与社会研究中的多数情况不同，在选举预测中，社会科学家必须对选举结果作出准确估计。选举结果本身即是预测准确与否的绝对标准。

（1）奥夫·兰登还是富兰克林·罗斯福

《文摘》杂志是美国一家大杂志。杂志编辑部在大选前用邮寄明信片询问投票倾向的办法正确预告了1920、1924、1928和1932年的当选总统。1936年，这家杂志根据电话簿和私人汽车登记簿上的名单寄出了一千万张询问卡，从调查情况看，共和党候选人奥夫·兰登以57%的优势超过仅占43%的民主

党候选人富兰克林·罗斯福。但选举结果表明罗斯福获胜连任总统。问题出在抽样方法上：在大萧条时期拥有汽车、电话的是富人，从电话簿和汽车登记簿上抽样导致样本中富人过多，而穷人基本上都为罗斯福新政投了赞成票。

（2）戴维还是杜鲁门

在1936年大选中崭露头角的民意测验者乔治·盖洛普正确地预言了罗斯福的当选。盖洛普的名字后来成为“民意”的同义语。他之所以作了成功的预测是由于运用了定额抽样的方法。定额抽样建立在了解抽样总体性质的基础之上，其中包括不同的性别、收入、年龄的人在总体中所占的比例。按这些比例分配定额：样本中应有几名贫困的、白种的、农村的男人，应有几名富裕的、黑种的、城市的女人，等等。定额是根据那些与研究关系最密切的变量确定的。盖洛普用这种方法保证了他的样本中各个阶层各种类型的人都占有正确的比例。

盖洛普及其美国民意测验所运用定额抽样的方法正确地预测了1936、1940和1944年大选结果。然而在1948年，盖洛普和其他许多民意测验机构都陷入了预测失败的窘境，他们预言纽约州长、共和党候选人汤玛斯·戴维将战胜民主党候选人杜鲁门。

预测失败的原因包括：第一，他们多在十月份就结束了调查，而到十一月大选这一个月间支持杜鲁门的趋势高涨。其次，许多选举人一直犹豫不定，在临投票时大批转向杜鲁门一方。然而最重要的一点是盖洛普样本的代表性出了偏差。定额抽样的方法是他以前成功的原因也是他这次失败的原因。到1948年，经过第二次世界大战，美国大量人口从农村流向城市，而盖洛普对人口总体的知识仍停留在1940年的状况。城市居民倾向于投民主党的票，因此，样本中过大的农村人口比例造成了对民主党投票人数的过低估计。

（3）两种基本抽样方法

在1948年，一些研究者开始采用概率抽样方法，结果比定额方法更成功。如今概率抽样成为社会科学研究抽取样本的基本方法。为了弄懂概率抽样的逻辑，有必要将其与非概率抽样相对照。本章主要论述概率抽样方法，对非概率抽样方法也略有涉及。

2 概率抽样的逻辑

（1）同质性与异质性

假如某个总体中的每一成员在一切方面都相同，抽样就没有必要了。对于这种极端同质性的情况，一个个体就足以成为研究整个总体性质的样本。例如研究人体的血细胞就用不着从许多人身上抽取血细胞的概率样本，从任何一个人身上抽的血样就足以进行研究了。

然而，当面对研究总体的异质性时则要求严格的抽样过程。现代抽样理论是由费舍（R·A·Fisher）开创的，他的名字至今还被用作某些普遍应用的统计方法的名称。总体是由不同的个体组成的。如果想通过从总体中抽取的样本描述总体，样本中必须包括同总体中一样的差异。概率抽样即为抽取充分反映总体差异的样本提供了有效的方法。

（2）有意识或无意识的抽样偏差

即使未受过特别训练，任何人都可以抽取一个调查样本。如抽取一个包括100名大学生的样本，可以径直到大学校园访问路过的学生。未受过训练

的研究者喜欢用这种抽样方法，但这种方法存在严重问题。

首先，研究者个人的好恶会影响抽样质量，导致样本缺乏对学生总体的代表性。假定研究者不喜欢具有嬉皮士外表的学生，担心他们会嘲笑自己的研究，于是就可能有意无意地避开这种人。即使研究者对走进图书馆的学生有意识地每隔十人选择一人，还是不能保证样本的代表性，因为这一样本中常去图书馆的学生将会比例过大。

（3）代表性和选择的概率

代表性一词并无精确的科学含义，但在关于抽样的讨论中是具有常识意义的有用概念。样本近似地代表总体，如总体中女性占50%，那么在一个有代表性的样本中女性也应当接近50%。

概率抽样的一个基本原则是：如果总体中的每一成员都具有被选入样本的同等机会，从这一总体中抽取的样本就能够代表总体。这种抽样方法称为概率抽样法。

概率抽样有两个特殊的优点：第一，它虽然不能完全精确地代表总体，但由于避免了偏差，可以比其他类型的样本对总体更具代表性。在实践中，概率样本比非概率样本对总体更具代表性。第二，

概率论可以对样本的精确性和代表性作出估计。

3 抽样的概念和术语

〔个体〕个体是收集信息的单位，它提供分析的基础。在调查研究中，个体常常是某种类型的人。其他类型的单位也可成为被调查的个体，例如家庭、社交俱乐部或公司等（注意：个体与分析单位在某一研究中常常是相同的）。

〔整体〕整体是研究中全部个体在理论上的假想的集合体。假如在某项研究中个体是单个的美国人，那么美国人就是整体。由于整体这一概念是超时空的，所以在研究中用处不大。

〔总体〕总体是在理论上明确定义的个体的集合体。美国人这个模糊的概念是调查的整体，而关于总体的描述将包括美国人的具体定义（如公民、居民及其他人口特征如老年、青年等）。将“成年的美国人”这一整体的概念转化为可用于调查的总体，需要明确规定“成年”的岁数定义和“纽约”的界线。

〔调查总体〕调查总体是研究者从中实际抽取调查样本的个体的集合体。总体是在理论上明确定义

的整体，但在实践中很难做到使符合这一理论定义的一切个体都有被选入样本的机会。甚至当掌握抽样总体的名单时，这个名单也往往是不完备的。有些学生由于注册过程中的技术性错误而没有进入学生花名册，有些人不愿意把自己的名字和电话号码登在电话簿上。研究者常常对调查总体进一步做严格的定义。美国全国民意测验由于某些特殊原因常常删去夏威夷和阿拉斯加。

〔抽样单位〕抽样单位是在抽样的不同层次选择个体的单位。在简单的、单层次样本中，抽样单位就是个体本身。然而在较为复杂的样本中则需要根据不同层次采用不同的抽样单位。例如，首先在某城市中抽取普查区的样本；然后在入选的普查区中抽取家庭的样本；最后在入选的家庭中抽取成年人的样本。在这三个层次上，抽样单位分别为普查区、家庭和成年人，亦称初级抽样单位，次级抽样单位和终级抽样单位。

〔抽样范围〕抽样范围是研究者从中抽取样本的抽样单位名单。如果一个简单样本是从学生花名册中抽取的，这个花名册就是抽样范围。如果一个复杂总体样本的初级抽样单位是普查区，全部普查区的名单就构成了抽样范围。

〔观察单位〕观察单位是我们据以收集信息的

个体或个体的集合体。分析单位和观察单位往往是相同的但也有例外。研究者可以通过访问家长（观察单位）收集关于家庭成员（分析单位）的信息。

〔变量〕变量是由一组相互不包容的取值组成的，如性别、年龄、就业状况等。通常调查的目的在于描述总体中某变量取值的分布。研究者可通过不同年龄的研究对象在总体中所占的比例描述其年龄分布。顾名思义，变量应是有差异的，如果某总体中所有的个体都只有一种取值，这个取值就是常数而不是变量的组成部分了。

〔参数值〕参数值是关于总体中某一变量的综合描述。例如，某城市中所有家庭的平均收入和这一城市人口的年龄分布就是参数值。研究的一项重要内容即在样本观察的基础上对总体参数值做出估计。

〔统计值〕统计值是关于调查样本中某一变量的综合描述。例如，从一个调查样本中计算所得的平均收入和样本的年龄分布即统计值。我们通常用样本统计值对总体参数值做出估计。

〔抽样误差〕在使用概率抽样方法时，样本统计值与总体参数值不可能完全一致。概率论使我们能够对某一样本设计可能导致的误差程度做出估计。后面还将详细讨论抽样误差问题。

[置信水平与置信区间]置信水平与置信区间是抽样误差估计的主要因素。我们用置信水平表达样本统计值落在某一指定区间的概率。例如可以说我们有95%的信心肯定样本统计值会落在总体参数值正负5%的两点之间。当某一统计值的置信区间增加时，我们的信心随之增加。例如我们有99.9%的信心肯定样本统计值会落在总体参数值正负7.5%的两点之间。

4 概率抽样理论和抽样分布

（1）概率抽样理论

调查抽样的最终目的在于通过对样本统计值的描述准确勾划出总体的面貌。概率抽样可以帮助我们实现这一目标并对其准确程度作出估计。

随机抽样是这一过程的关键。在随机抽样过程中，总体中的每一个体都有同等的机会入选，且每一事件都是相互独立的。随机过程最常用的例子是投掷硬币，每次投掷的结果（正面或反面）都是相互独立的。

采用随机抽样方法的原因是双重的。一方面，它可以检验出抽样偏差；另一方面，它是概率论的

重要内容，而概率论是总体参数估计和误差估计的基础。

（2）十个个案的抽样分布

我们用一个简单例子来介绍概率抽样的方法。假定有序号分别为0－9的十个人，每人手中美元数与序号同，求每人平均拥有美元数。首先求十人的钱数总和（45美元），然后得到平均数为4.5美元。

如果我们不看总体而用随机样本来估计平均值。首先用容量为1的样本来估计平均值，这一估计可能为0美元到9美元（图7－1）。图中的十个点代表我们得到的十个“样本”平均值，它们都是总体平均值的估计值。点子在图中的分布称为抽样分

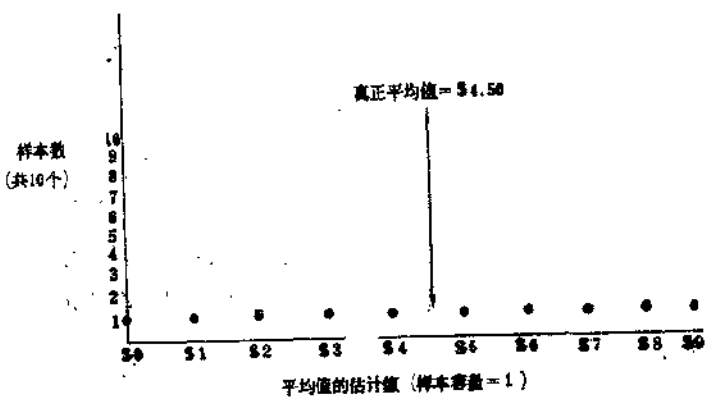


图7—1　容量为1的样本之抽样分布

布。如果样本容量为2，我们的估计将会更准确一些（图7—2）。我们将有45个样本：0美元和1美元，0美元和2美元，……，7美元和8美元，8美元和9美元。其中有些样本会产生相同的平均值，如0美元和6美元，1美元和5美元，2美元和4美元，这三个

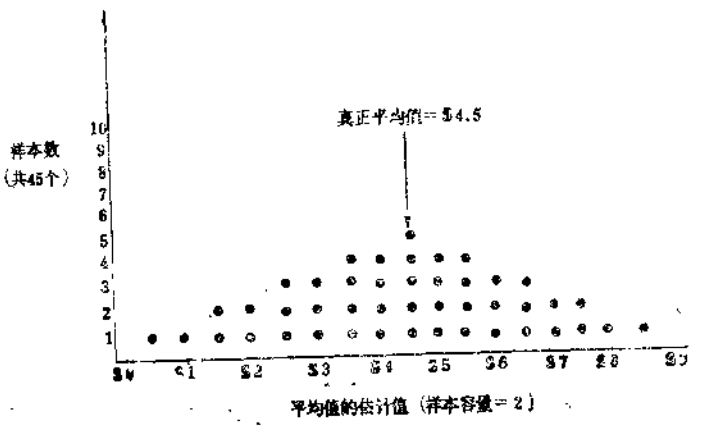


图7—2　容量为2的样本之抽样分布

样本的平均值都是3美元。图7—2中3美元一列的三个点即这三个样本的平均值。

如果样本容量继续增大，平均值的估计值会发生什么变化呢？图7—3是样本容量分别为3、4、5、6的抽样分布。样本容量的增大提高了平均值估计值的准确性。

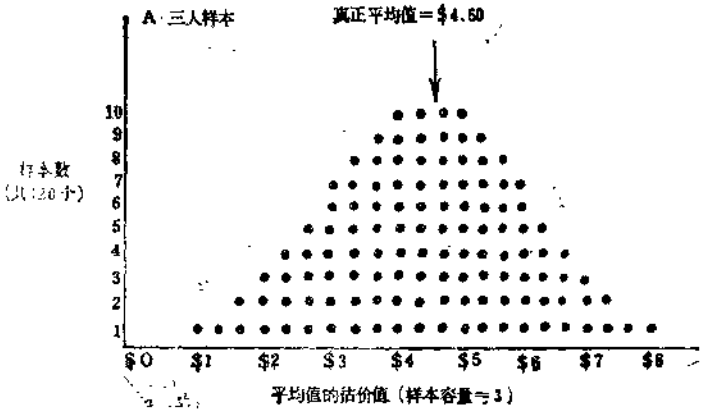


图7—3　容量为3、4、5、6的样本之抽样分布

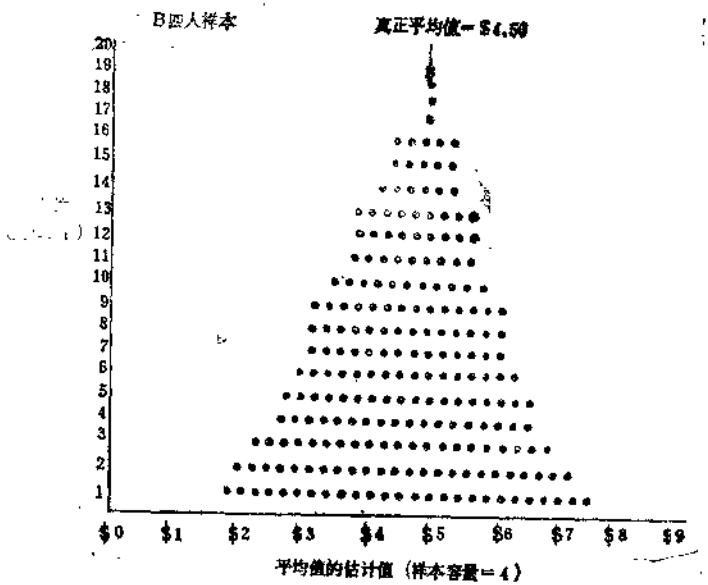


图7—3　（续）

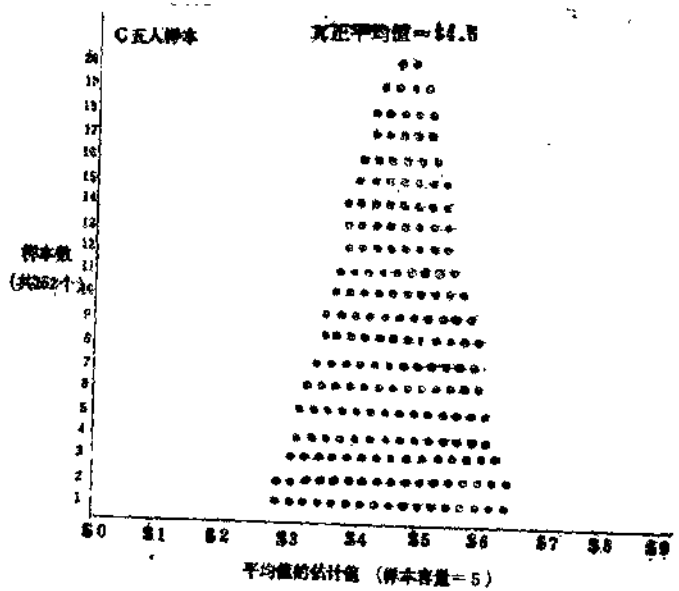


图7—3 （续）

（3）二项抽样分布

假定我们想研究某大学学生对学生守则的看法。

这个大学有二万多名学生，我们在学生花名册上根据随机数字表抽取100人的样本。调查结果表明，48名学生赞成守则，52名反对。我们再抽取另一个100人的样本，结果是51人赞成，第三个样本是52人赞成。图7—4代表了三个不同样本中赞成学

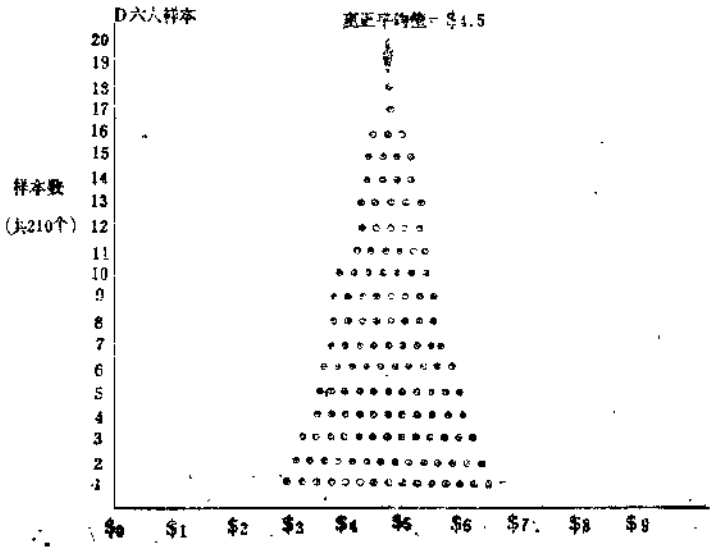


图7—3 （续）

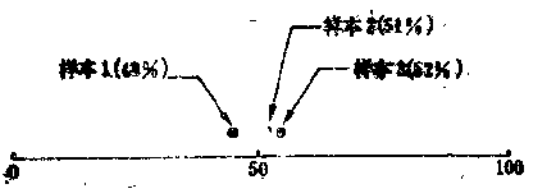


图7—4 三个假想的调查结果

生守则的人数。如果我们取100个样本，抽样分布就会呈现出图7—5的情形。概率论关于这种抽样分

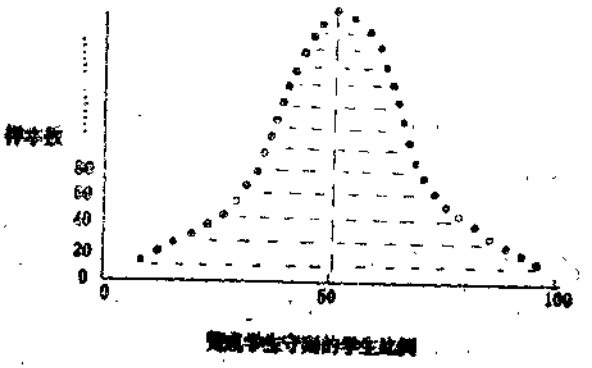


图7—5 抽样分布

布有几项重要规则。

第一，如果从总体中抽取许多相互独立的随机样本，这些样本提供的样本统计值将以一种可知的方式分布于总体参数值周围。从图7—5可以看出估计值集中在50%近旁。第二，概率论为我们提供了一个估计样本统计值与总体参数值接近程度——标准差S的公式。

概率论表明，接近34%（0.3413）的样本统计值将落在高于总体参数值一个标准差的范围内，另外34%落在低于总体参数值一个标准差的范围内。大约95%的样本统计值将落在总体参数值正负两个标准差范围内；99.9%的样本统计值将落在总体参数值正负三个标准差范围内。标准差与总体参数值

成正比关系，与样本容量成反比关系。我们就是在这个意义上讲置信水平的：样本估计值落在总体参数值一个标准差范围内的概率为68%；落在两个标准差范围内的概率为95%；落在三个标准差范围内的概率为99.9%。

我们常常并不知道总体参数值，因此用样本统计值对总体参数值做出估计和推论。这种估计和推论会不可避免地存在一定程度的误差。为了了解赞成学生守则的学生比例，我们抽取了一个100人的随机样本。调查结果表明，最佳估计为50%且赞成者占40%到60%（正负两个标准差）的概率为95%。从40%到60这一区域就被称为置信区间。关于抽样的结果的准确说明必须包括置信水平和置信区间这两项内容。

在这一讨论中，我们只涉及了一种类型的统计值：由二项变量产生的百分比。对其他类型的统计值的检验也应采用相同的逻辑，例如平均收入等。由于平均值的计算更复杂一些，本节只选择了二项分布来介绍抽样的基本逻辑。

5 抽样设计之类型

（1）简单随机抽样

简单随机抽样是调查统计中的基本抽样方法。研究者为抽样名单上的每一个体编码，然后根据随机数字表取样。如果抽样名单可以输入计算机，计算机将会自动抽取简单随机样本。

（2）系统抽样

在实践中人们很少运用简单随机抽样的方法，因为它十分费事，效率不高。简单随机抽样要求调查总体中全部个体的名单，而当人们得到这一名单之后往往更愿意用系统抽样的方法。

系统抽样方法是从名单上每隔一定的数抽一个样。例如，名单上有一万个个体，样本容量为一千，则每十人选出一人。为了避免偏差，起始数字必须是随机决定的。因此这种方法又称为随机起始的系统抽样方法。在系统抽样中有两个常用的术语，一个是抽样间距，指两个被抽出的个体之间的标准距离，如前例中的抽样间距为10。另一个是抽样比例，指从总体中抽出的个体所占的比例，如前

例中的抽样比例是1/10。系统抽样与简单随机抽样的结果基本相同，有时前者甚至比后者更准确。

但是运用系统抽样方法有一个潜在的危险，即当名单的排列存在周期性时，例如，我们从军队花名册上每十人取一人，而名单恰恰是按十人一班登记的，结果被抽中的全是班长，或一个班长也未被抽中。再举一例。假定我们在一幢公寓大楼的单元中抽样。单元号码为101，102，103，104，201，202，……。万一抽样间距与单元号码的周期一致，抽样结果则有可能全是西北角的单元或靠电梯的单元。假如这种单元有共同特点（如房租较高），样本就会有系统偏差。为避免这种偏差必须仔细研究系统抽样名单，看它是否存在会造成偏差的特别顺序。如果发现这种情况，可在存在周期性的部分采用简单随机抽样的方法抽样。

（3）分层抽样

简单随机抽样和系统抽样在一定程度上保证了样本对总体的代表性同时包含一定程度的抽样误差。运用分层抽样的方法则可得到更大程度的代表性同时减少抽样误差。

影响抽样误差的因素有二：第一，大样本比小样本抽样误差小；第二，同质性总体比异质性总体

抽样误差小。分层抽样即以第二个因素为基础，不是在大总体中直接抽样，而是按照总体中各个同质层所占的比重抽样。例如在大学生中进行分层抽样，可以按年级、性别和平均分分层。用这种方法可以保证样本中包括适当数量的平均分为甲等的一年级男生、平均分为甲等的一年级女生，依此类推。

分层的最终目的在于把总体分成一个个同质层，（层与层之间则是异质的），给每个层次分配适当的名额。根据哪一变量分层通常取决于哪些变量是已有层次的。教授名单常常是按系分好的，政府文件常常是按地区分好的。除此之外，还可根据研究中要求精确性较高的变量来分层。由于性别与许多事相关，可用性别分层，教育程度和地理位置也可作为分层的根据。

根据每层在总体中所占的不同比例分配好名额以后，用简单随机抽样或系统抽样的方法在各个层次内抽样。假如平均分为甲等的一年级男生占学生总数的百分之一，样本容量为1000，就应当在平均分为甲等的一年级男生中抽一个容量为10的样本。

6 举例：在大学生中抽样

本节介绍一个实际的抽样设计，这是1968年在夏威夷大学所做的一个抽样。

（1）调查总体与抽样范围

抽样从学校行政部门保存的学生注册记录开始，记录中包括学生的姓名、地址、社会安全号码、专业、班级、年龄、性别等信息。并非所有的学生都与研究目的相符，因此应当在更严格的意义上定义调查总体。经过考虑，最后把调查总体定义为15,225名1968年秋季注册的在校生，其中包括各系各班，本科生和研究生，美国学生和外国学生。计算机抽样范围限制在符合上述定义的学生之内。

（2）分　层

研究者决定第一步只按学校班级分层，以后如有必要再在班级内进一步按性别、专业等分层。

（3）抽　样

根据分班名单进行系统抽样，样本容量初步定

为1,100，即抽样比例为1/14。首先在1到14之间确定一个随机数字，从这个学生开始每14名选出1名。入选名单确定后，按这些人的地址寄出问卷。

（4）修改样本

在问卷寄出之前研究者发现由于意外开支经费略嫌不足，于是决定用系统选择（随机起始）的方法删去样本中三分之一的人，结果样本容量被修改为770人。提到这项修改旨在说明，在整个研究过程中，计划可能需要随时修正。剩下的770人仍旧可以代表调查总体，虽然样本容量的缩小无疑会使抽样误差相应增加。

7 多段类集抽样

前面谈到的是根据名单抽样的简单程序。遗憾的是在许多研究中很难搞到全体名单。例如某城市、某省或全国居民，美国的全体大学生等。遇到这种情况，抽样设计就大大复杂化了。此类设计一般是首先抽取类集，然后在类集内再作个体抽样。

例如在某城市抽样，我们没有该城全体居民的

名单，但有各个街区的名单。可以先从全部街区中抽取样本，然后设法搞到每一入选街区的居民名单，在这些个人中继续抽样。还有一种更加复杂的设计，先在街区中抽样，列出每个入选街区的家庭名单，在这些家庭中抽样，然后列出每个入选家庭的个人名单，再从这些个人名单中抽样。

多段类集抽样包括两个基本步骤的反复运用：开列名单和抽取样本。这种方法效率很高，缺点是不够精确。简单随机抽样只造成一个抽样误差，而两段类集抽样就会造成两个抽样误差。如在第一阶段被抽中的富裕街区超过比例，接下去被抽中的富裕家庭又超过比例，于是样本就会与总体相去甚远。解决这一问题的最佳方法在于对各段抽样数量的考虑。

例如在某城市抽取2,000人的样本有几种不同的抽法。其中极端的抽法有两种：抽取1个类集，然后从中抽取2,000个个体；或抽取2,000个类集，每个类集中抽取1个个体。当然这两种方法都不好。两极之间的最佳抽样方法是什么？抽样分布的逻辑为我们提供了一般原则。

前面讲过，有两个面素可以减少抽样误差：一个是增加样本容量，另一个是增加个体的同质性。因此：抽取的类集或个体都是越多越好。可是由于

样本容量已定，类集多则个体少，反之亦然。这时，对同质性因素的考虑可以使抽样设计摆脱这个两难窘境。

组成类集的个体通常具有同质性，某一特定街区的居民通常比城市中其他居民有更多的共同点。因此个体可以相对少些，类集则相对多些。但如果类集过多又会降低效率（每个类集都需要一个总名单）这将违反我们的初衷。最终的选择取决于这二者之间的平衡。按照一般作法，假如样本总数为2,000个家庭，就抽取400个街区，每个街区5家。

（1）多段类集抽样中的分层

分层的技术可用于抽样的每一阶段。类集也可以象个体一样分出层次。例如要在全国的教堂中抽样，可将教堂先按地区、规模、在城市还是在农村等分层。在每一层次中运用简单随机抽样或系统抽样方法抽取样本。在同质层次中抽样可以减少抽样误差，因此应当在抽样的每一阶段都运用分层技术。

（2）规模比例概率抽样

规模比例概率抽样是另一种用于大规模研究项目的更加复杂的类集抽样方法。假如某城市有1,000

个街区，我们先从中抽100个，则每个街区被抽中的机会为100/1000，即1/10。然后在那些入选街区中每10户选1户，入选机会亦为1/10。这样，每户在全部户数中被抽中的机会为$1/10 \times 1/10 = 1/100$。由于每户都有同等机会被选中，这样选出的样本就能代表这一城市中的全部家庭。

但是，一个潜在的问题是：街区大小不一。假设城市中一半人口集中于10个街区的高层建筑中，另一半人口分散在900个街区的独立式房屋中。在第一步抽取1/10街区时，那10个人口集中的街区可能一个也未入选。当各个类集大小相差太多时就应当采用修正的抽样设计，所谓规模比例概率抽样。这种方法可以防止上述问题发生，从而产生一个每一个体都有同等机会入选的样本。

顾名思义，规模比例概率抽样是根据每个类集的规模来分配样本名额的。因此，在样本中，200户的街区所占名额就应当二倍于100户的街区。例如，甲家所在的街区有100户，乙家所在的街区只有25户，已经决定从每个街区选5家。如果甲家所在的街区入选，甲家入选机会为5/100；如果乙家所在的街区入选，乙家入选机会为5/25或20/100。于是在抽样的第二阶段，乙家入选的机会是甲家的四倍。

运用规模比例概率抽样方法选择个体的总概率，应按如下方法计算：

（1）入选类集的概率应当等于其全部个体在总体中所占的比例乘以入选类集的数量。

（2）每一类集中入选个体的概率应当等于每一类集内入选个体的数量除以该类集中全部个体的数量。

（3）每一个体入选的总概率等于（1）乘以（2）。

例如某城市有2,000个街区和100,000个家庭。我们计划抽样1,000户。这就是说，每户入选机会应为1,000/100,000，即0.01。我们决定抽出200个街区，每个街区选出5户。

假定某个街区有100户人家，该街区入选概率应为

$$200(\text{入选街区数}) \times \frac{100(\text{该街区户数})}{100{,}000(\text{城市总户数})} = 0.2$$

假如该街区入选，每户在第二阶段入选的概率为

$$\frac{5(\text{每街区入选户数})}{100(\text{该街区的户数})} = 0.05$$

0.2乘以0.05即入选总概率（0.01），

假定某街区只有20户人家，该街区入选的概率应当只有200×20/100,000即0.04。假如该街区入

选，每户入选机会为5/20即0.25。入选总概率为0.04乘以0.25，即0.01。

上述两例中个体入选的概率是相等的：

例一：$200 \times \frac{100}{100,000} \times \frac{5}{100} = 0.01$

例二：$200 \times \frac{20}{100,000} \times \frac{5}{20} = 0.01$

两例唯一区别之处在于街区的户数，让这个数字作一次分子再作一次分母，它们就互相抵销了。无论街区的大小多么不同，各户入选的概率是相同的。

（3）非比例抽样与加权

当所有的样本个体都有同等机会入选时，每一个体的比重为1，但在某些研究中需要运用非比例抽样与加权的方法。非比例抽样是指从总体的各部分中按不同的比例抽样，以保证每一部分都有足够的个案供研究用。例如，某城市中郊区人口仅占四分之一，但由于研究者对郊区很感兴趣，不愿使自己的样本中郊区所占比例太小，于是决定城区和郊区在样本中各占一半。这种抽样方法使郊区家庭入选的机会高于应有比例。

为了遵循样本个体入选机会相等的原则，必须再用加权的方法，使郊区的比重仅占1/4，城区的比重占3/4。这一加权程序可简化为将在郊区取的

样以1为1，而城市取的样以1为3，这样在数据分析阶段郊区与城区就又恢复了各自应占的比例。

（4）加权的精确度

加权的精确度可精确到小数点后几位，也可粗略一些。如前例中提到的郊区人口占1/4就不可能是整整1/4，可能是总体的0.25001，0.26或0.2816。加权时可用这些精确数字也可就用0.25，依研究本身对精确度的要求而定。如果某项研究可以容许少量误差，就不用浪费时间和精力去做精确加权。

8 举例：在女教徒中抽样

某项研究的目的是了解某个教区内女教徒的观点。首先需要抽取一个可以代表这一教区内全体女教徒的样本。由于没有全体女教徒的名单，样本设计为多段类集抽样。在抽样的第一阶段，用规模比例概率抽样方法在教堂中抽样，然后进一步在入选教堂的女教徒中抽样。

（1）在教堂中抽样

这个教区有大小不同的100多个教堂，样本容

量为500人，运用规模比例概率的方法抽取25个教堂，再从每个入选教堂抽取20个女教徒。抽样的第一个步骤是为全部教堂造表（如表7—1），

表7—1 教堂名册

教堂	教徒人数	累积教徒人数
教堂甲	3,000	3,000
教堂乙	5,000	8,000
教堂丙	1,000	9,000
⋮		

造表结果表明，全教区教徒人数约为200,000人。用系统抽样的方法抽取25个教堂，抽样间距应为8,000人（200,000/25），随机起始。假定这个随机数字是4,538，由于这一数字落在教堂乙的范围内（3,001—8,000）教堂乙就入选了。接下去每隔8,000人取一个教堂，从而使教区中的每个教堂都按照其规模得到适当比例的入选机会。拥有4,000教徒的教堂的入选机会二倍于拥有2,000教徒的教堂，十倍于拥有400教徒的教堂。

（2）在女教徒中抽样

选出25个教堂后，应设法得到这些教堂中女教徒的名单。抽样间距决定于该教堂女教徒总数和抽样人数（20）。假如某教堂有2,000女教徒，则抽

样间距为100。

这一抽样设计建立在各教堂男女教徒比例相等的假设之上。事实上有些教堂女教徒多于男教徒，另一些教堂女教徒少于男教徒。考虑到这一研究的特殊目的，可以忽略这种轻微的不平衡。

（8）概率抽样综述

本章讨论了调查研究的主要抽样方法——概率抽样方法，即从总体中按照已知的非零概率随机抽取样本。这种方法之所以成为抽样的最佳方法原因有二：

第一，概率抽样避免了研究者在抽样过程中有意或无意的偏差。假如某总体中的所有个体都有同等机会被抽入样本，则样本会相当准确地代表总体。

第二，概率抽样可以对抽样误差做出估计。虽然概率样本不会与总体完全一致，但这种方法使研究者可对期望的误差做出估计。

9 非概率抽样

有时，情况不允许做概率抽样，如我们想研究

贪污者时，没有也不可能搞到一个包括全体贪污者的名单。在另外一些情况下，做概率抽样是可能的，但是不适当。于是就需要运用非概率抽样的方法。

（1）有目的、有判断的抽样

有少数研究应当在理解总体与个体的基础上抽样，或以研究目的为基础抽样。尤其是在开始设计问卷的时候，研究者可能想选择一些观点差异悬殊的人来检验问题的设计是否恰当。虽然这种调查结果不能代表任何有意义的总体，但会有效地发现问卷设计中的缺失。实际上这种作法不能算是做研究，只是一种防止偏差的措施。

有时，由于统计某一总体成员的数量几乎不可能，研究者只好退而研究这一大总体中的某个小部分。例如研究者想研究某学生运动的领袖们。许多领袖可以很容易被辨认出来，但是要定义所有的领袖并在他们当中抽样几乎是不可能的。这时只能采用在最明显的学生领袖中抽样的方法。再如研究者想将左翼学生与右翼学生加以比较。由于不可能计算他们的人数并从中抽样，他可能决定仅在“争取民主社会学生联合会”和“争取自由青年美国人”这两个组织的成员中抽样。虽然这种样本设计不能描述

全体左翼学生与右翼学生的情况，但对一般的比较研究来说已经足够了。

（2）定额抽样

前文已提到过，定额抽样是使盖洛普在1936年避免灾难又在1948年给他带来灾难的方法。定额抽样与概率抽样都以代表总体为目的，但达到这一目的的方式却很不同。

定额抽样从对总体性质的了解开始，如在某一总体中不同性别、年龄、教育程度、种族的人各占多大比例等。然后从具有上述属性的人中搜集数据，并按各类人在总体中的比例赋予他的适当的比重。如此收集来的数据从理论上讲应当能够代表总体。

定额抽样方法的固有问题在于：定额的比例必须精确，但最新的关于总体性质变化的信息并不容易得到，这就是盖洛普1948年失败的原因。此外还有抽样中的偏差。

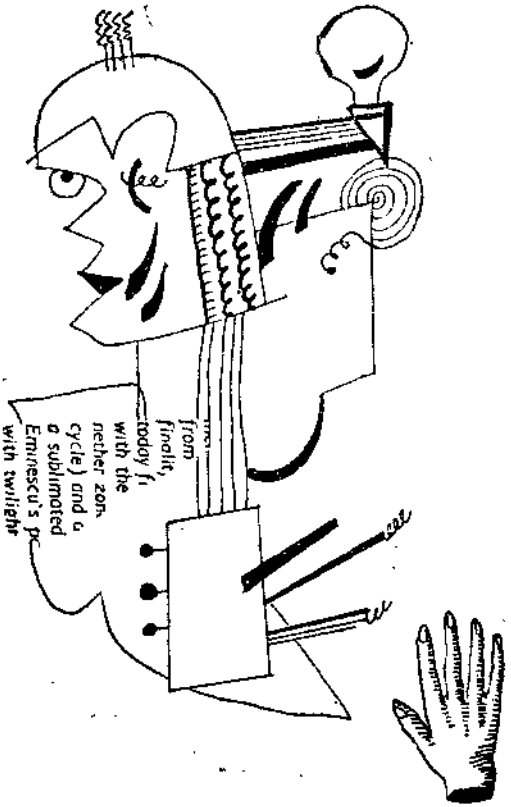
（3）依靠现成研究对象

依靠现成研究对象是指在街头或类似的公共场所拦住行人做调查。虽然这种作法很常见，但它绝不是一种科学的合格抽样方法。大学中的研究者常

常在学生的大课上搞调查，这种作法被普遍采用仅仅由于其容易和省钱的特点，这样收集来的数据没有什么价值。这一方法也许可以用来检验问卷设计的好坏，但对全体学生绝无代表性可言。

某项关于医学院学生和家庭医生对营养与癌症方面知识程度的研究是这样抽样的：这一研究的学生总体由明尼苏达医学院四年级学生组成；医生总体由参加明尼苏达大学医学进修班主办的“家庭医疗的过去与现在”课程的医生组成。这个研究有什么代表性？它既不能提供全美甚至也不能提供明尼苏达州的医科学生与家庭医生的情况比较。这些来上课的医生是些什么人？我们可以推测，他们可能比其他医生更关心医学进修，但即使这一点也不能肯定。结果是我们不知道这样的研究结果究竟有什么用处。

第三篇　观察方式


from
finalit,
today fi
with the
nether zon
cycle) and a
a sublimated
Eminescu's pc
with twilight

第八章
实——验

本章讨论一种与一般科学研究方法最为相似的社会研究方法：以实验作为科学观察方式。实验一般包括：(1)实行某项措施(2)观察这一措施的效果。社会科学研究者一般也是选择一组研究对象，对他们实行某项措施，然后观察后果。本章将讨论社会科学实验的逻辑和技巧。

1 适于采用实验方法的研究课题

实验的方法特别适用于概念和命题相对有限的、定义明确的研究项目以及假设检验。在以解释或描述为目的这两类研究中，它更适用于解释性研究。假定我们想研究对黑人的种族歧视及克服这种偏见的途径。我们首先假设了解黑人在美国历史中

的贡献将会减少对黑人的偏见，然后对这一假设进行实验的检验。在实验的开始，我们先了解一组实验对象的种族偏见程度，然后给他们放映一部描述黑人在科学、文学、政治和社会发展上对国家作出的贡献的纪录影片。最后，通过度量实验对象种族偏见程度的变化确定放映这部电影是否起到了降低种族偏见程度的作用。

实验的方法还被大量运用于小群互动的研究，且相当成功。例如，交给实验对象小组一项任务，如对某事作出评价。然后观察这个小组是以何种方式组织起来完成这一任务的。在多个类似实验中，可设计不同的任务和报酬，通过观察各个小组在这些不同条件下如何行动，我们可以了解到大量有关小群互动的信息。

一般认为实验就应当在实验室进行。确实有一部分实验是如此，但不是全部。社会科学家经常进行所谓“自然性实验”，即在社会事件的一般过程中进行的“实验”。

2 古典实验方法

无论是在自然科学还是社会科学研究中，多数

传统形式的实验包括三对主要内容：(1)自变量与因变量，(2)实验组与控制组，(3)事前测验与事后测验。

（1）自变量与因变量

实验的基本内容是检验自变量对因变量的影响。通常自变量是实验中的刺激因素，且多为二项变量（只有两个取值的变量，如刺激因素存在与刺激因素不存在）。前述关于种族歧视的例子中，偏见程度是因变量，放映有关黑人历史的影片这一措施是自变量。研究者的假设是偏见来源于缺少关于黑人历史的知识。实验的目的是检验这一假设的正确性。

可用于实验方法的自变量和因变量是无穷无尽的，但应提请人们注意的一点是，某项实验中的自变量可能成为另一实验中的因变量。例如偏见程度在前例中是因变量，但在检验偏见对选举行为的影响的实验中则是自变量。

（2）事前测验与事后测验

在最简单的实验设计中，我们先用因变量来度量研究对象（事前测验），经过实施自变量这一刺激因素，再用因变量来重新度量研究对象（事后测

验）。因变量在前后两次度量中的区别则被归因于自变量的影响。

如在关于种族歧视的例子中，应当从对实验对象的偏见程度进行事前测验开始，用问卷询问对黑人的态度，并度量每个对象的偏见程度和全组平均偏见程度。在放映关于黑人历史的影片后，再用同一问卷进行事后测验，以度量个人和小组偏见程度的变化。如果我们发现事后测验中偏见程度有所降低，则可能得到对研究对象放映关于黑人历史的影片确实降低了偏见程度的结论。

在此类研究中我们面临着一个特殊问题：即使研究对象的态度未改变，他们第二次的回答也可能同第一次不同。因为在事前测验时他们不明白研究的目的，而在事后测验时他们可能已经发现研究者的意图是度量其偏见程度。由于人们都不愿被视为有偏见者，因此他们改变了答案。于是，仅从结果上看电影降低了偏见程度实际却并未起到这种作用。这是社会科学研究中的一般性问题：研究本身会改变研究对象。本章其他部分还将讨论解决这一问题的办法。

（3）实验组与控制组

防止研究本身影响研究对象的主要方法是设置

控制组。一般实验室实验也很少只用一个实验组，而是设置一个对其不施行实验刺激的控制组。

仍以种族歧视的实验为例。首先要求两组成员填写同样问卷进行事前测验，然后只给一个组（实验组）放映电影，最后在两组中同时进行事后测验。图8-1即这种基本实验设计的图示。

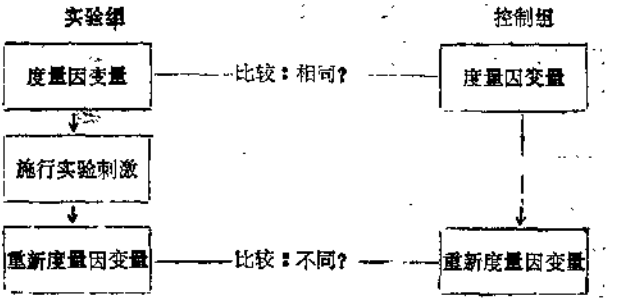


图8—1 基本实验设计图示

设置控制组可以使研究者控制实验本身对研究对象的影响。如果参加实验这件事本身会使偏见程度降低，则两组都应降低。如果事后测验结果表明两个组的偏见程度都降低了，则说明影响在于实验本身或其他外部原因，而不是由于放映电影。假如只有实验组降低了，或两组都有所降低但实验组降低的幅度更大，则说明放映电影对降低偏见程度是有作用的。

社会研究中设置控制组的必要性是因罗易斯里

斯伯格（F.J.Roethlisberger）和迪克逊（W.J. Dickson）在二十年代末三十年代初的研究而引起注意的。他们的研究旨在发现提高工人劳动生产率的办法。在研究过程中他们发现某些工作条件的改善，如增加照明设备可以提高工人的劳动生产率。当照明继续增加时，劳动生产率继续上升。为了从反面证实这一发现，他们减低了照明度，期望劳动生产率会相应下降，但令人惊讶的是劳动生产率仍是上升！事实证明，实验期间劳动生产率的上升是由于工人们发现有人在观察他们，而不是由于劳动条件的改善。这一现象被命名为豪森效应。从这一研究开始，社会研究者才懂得要谨慎地对待实验本身可能带来的影响。

3 研究对象之选择

（1）概率抽样

概率抽样的逻辑与技巧可以提供选择情况相似的两组人的方法。研究者从同一抽样范围内抽取两个随机样本，如果这两个随机样本都分别与总体相似，那么它们本身也应当相似。然而情况并非如此简单。由于样本容量对随机样本的代表性影响甚

大，具体地说，容量小于100的概率样本就不是很有代表性了，而社会科学实验又很少用大样本，因此概率抽样的方法很少用于实验。

（2）随机过程

实验者可以随机地将一大组研究对象分为实验组和控制组。具体方法包括用随机数字表分组或按单双数分组。假如实验者一共征集到40名研究对象，并将其随机地分为两组，可以肯定这两组的情况会十分相似。当然，研究对象的数量是重要的，如果只有两个研究对象，尽管将其随机地分为实验组和控制组，二者情况也不会是相似的。总之，研究对象数量越多，随机过程的作用就越大。

（3）匹配过程

有时，实验组和控制组的可比性可通过直接的匹配过程，其方法与定额抽样近似。假如研究对象中有12名年轻白人，则可以随机地选出6名为实验组成员，其他6名为控制组成员。假如研究对象中有14名中年黑种女人，则每组7名。最有效率的方法是制作定额表（见图8-2）。最理想的情况是各栏均为双数，因此可以刚好分给实验组和控制组各一半人。

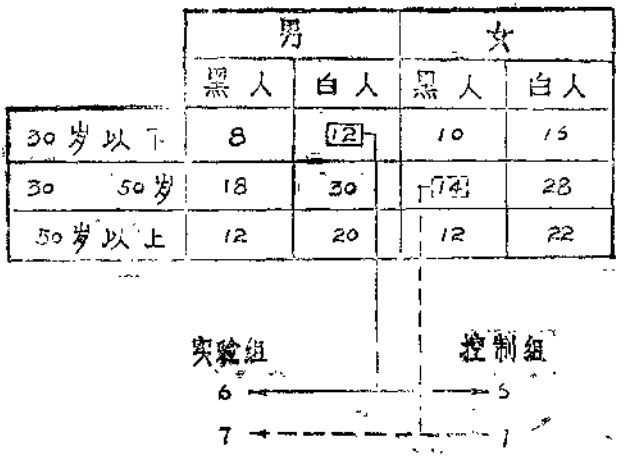


图8—2 定额表举例

匹配的另一具体作法是，搜集多于要求数量的研究对象，一一检查之，一旦发现一对类型相似的研究对象，即将其随机分到两个组中，放弃那些配不上对的研究对象。

无论用什么方法，目的是使两组的情况大致相似。如，它们应当有大致相同的平均年龄、性别比例、人种比例等等。应当使研究对象在哪些变量上相匹配依研究目的而定。一般来说，应当使他们在那些与因变量有关的变量上相匹配。如研究偏见程度应使两组在教育程度、民族和年龄上相似。此外，应当以事前测验填写问卷的情况为依据，使两

组在偏见程度上相匹配。

4 实验设计的不断改进

多那德·坎波贝尔（Donald Campbell）和朱莉安·斯坦雷（Julian Stanley）在一本关于研究设计的书中介绍了16种实验和准实验设计方法。其中有三种原始的方法曾被大量应用于专业性较差的研究中。第一种设计称为一次性个案研究，具体作法是在施行实验刺激之后度量一组研究对象在某一因变量上的变化。如给一组研究对象放映关于黑人历史的影片，然后让他们填写问卷用以度量其种族偏见程度。

第二种设计加上了事前测验但不设控制组。这种单组事前-事后测验设计的问题在于难以发现自变量以外的其他因素的影响。

第三种设计有实验组和控制组但无事前测验。这种设计被称为静态组间比较。这种研究设计的缺点在于事后测验中即使发现实验组偏见程度较低也不能做出任何有意义的结论，因为不了解实验开始时两组情况是否相同，或许实验组的偏见程度原来就低于控制组。

图8-3是这三种研究设计的示意图，应充分注意这几种设计的缺点。

（1）导致实验内部失真的原因

内部失真是指从实验结果中所作的结论没有正确地反映实现过程。坎波贝尔和斯坦雷将导致实验内部失真的原因概括如下：

①历史性事件。在实验过程中发生的历史性事件可能影响实验结果。如在实验期间某位黑人领袖遇刺可能导致种族偏见程度的大幅度降低。

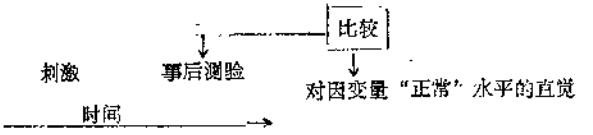
②人们的成熟。人们在不断变化，而这些变化会影响实验结果。在一个时间持续很长的研究中，人们会变老（也许会因此变得聪明一些？）；在短期研究中，人们也可能变得疲倦、饥饿、厌烦，而这些变化将会影响实验结果。

③实验本身的影响。实验过程本身会影响人们的行为，从而影响实验结果。人们会看出研究者的意图而试图把自己表现得比实际情况好些，回答研究者愿意听的话。

④不同的度量尺度。如果事前测验与事后测验的度量尺度不同，也会影响实验结果。也许事后测验中呈现出来的偏见程度的降低仅仅因为事前测验的度量尺度比事后测量的更加敏感一些。

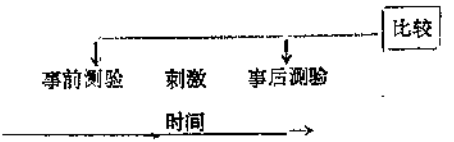
1. 一次性个案研究

对一个组施行实验刺激，度量其因变量，对事后测验结果是"高"还是"低"做直觉的判断。



2. 单组事前—事后测验设计

在一个组内度量其因变量，施行实验刺激，然后重新度量因变量，比较事前测验与事后测验的结果。



3. 静态组间比较

对一个组（实验组）施行实验刺激，然后将实验组与控制组在因变量上加以比较。

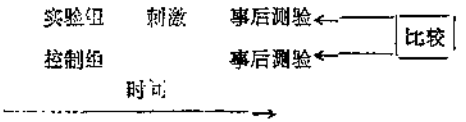


图8—3 三种不完善的实验研究设计

⑤向平均值的回归。以因变量极端值为基础的研究容易犯一种错误，称为向平均值的回归。例如，如果以一组数学最差的人为对象检验某种新教学法的功效，研究对象数学成绩的提高往往不是由于新教学法的作用，而是因为它已经低得不能再低了。统计学家发现，在极端点上，一组高人的子女的平均身高总是比父母的要低，一组矮人的子女的平

均身高总是比父母的要高。这一现象就是所谓“向平均值的回归”。危险在于把这种从极端点上发生的向心移动错误地归功于实验刺激的作用。

⑥选择的偏差。前面谈论过划分实验组与控制组时容易产生的偏差。组与组之间必须是可比的，否则比较将没有任何意义。

⑦实验夭折。有时由于研究对象在实验完成之前脱离实验，也会影响实验结果。例如实验组有些偏执分子看了关于黑人历史的电影愤而退出实验，致使事后测验中偏见“降低”的结果不可信。

⑧多种交互作用。有时多种原因的交互作用会影响实验结果。例如实验组比控制组的人厌倦临界值低，或实验刺激本身使人厌倦，因此在事后测验时实验组比控制组厌倦情绪高得多，从而影响实验结果。

⑨措施的扩散或模仿。在实验组和控制组对象并未完全隔离的情况下，前者可能把实验刺激的内容传达给后者。如实验组的人可能把黑人历史影片的内容告诉控制组的人，于是控制组等于也受到某种程度的实验刺激，而并未真正受到控制。

⑩补偿心理。在现实生活中所作的实验，例如新教学方法或新治疗方法的实验，可能会给实验组的对象带来实际利益，反之使控制组的对象蒙受损

失，于是产生了补偿心理的问题。如，医院工作人员可能由于感到对不起控制组的病人而给予额外的“感情关怀”。在这种情况下，控制组就不再是真正的控制组了。

⑪补偿竞争。在现实生活中进行的实验里控制组的对象有可能因为失去接受实验刺激的机会而努力去补偿这种损失。如实验组接受了一种新的数学教学法的刺激，而控制组将特别努力学数学想胜过那些“特殊”的实验对象。

⑫情绪低落。另一种可能是控制组由于感到受了损失而放弃努力。在新教学法实验中，情绪低落的控制组成员可能停止学习，捣乱或愤愤不平。

意识到这些问题后，实验者设法在设计中避免这些问题。因此古典实验设计更趋完善(见图8-4)。

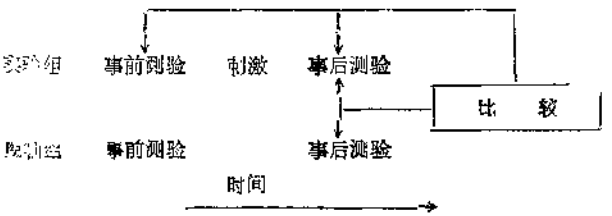


图8—4　古典实验设计

(2)导致实验外部失真的原因

除了内部失真之外，实验者还会遇到外部失真

的问题，即实验结果对于现实世界概括性较差的问题。即使实验结果是实验本身的精确反映，它是否真能反映客观世界呢？

坎波贝尔和斯坦雷对导致实验外部失真的原因也做了概括。危害实验结果概括性的原因之一是实验本身与实验刺激之间的相互作用。例如，我们发现在实验中影片的放映确实使种族偏见程度有所降低，但如果在电影院或电视台放映这部影片会产生同样的效果吗？仅包括实验组和控制组的古典实验设计难以解决这一问题，于是我们采用所罗门四组设计（如图8-5）来解决这一问题。

这种实验设计可做四种比较。假如关于黑人历史的影片确实降低了种族偏见的程度，既无内部失真也无外部失真，我们应当得到下列四种偏见程度变化的结果：

①第一组事后测验应低于第一组事前测验。

②第一组事后测验应低于第二组事后测验。

③第三组事后测验应低于第二组事前测验。

④第三组事后测验应低于第四组事后测验。

第三种和第四种结果可以避免实验本身与实验刺激间相互作用导致的外部失真现象。

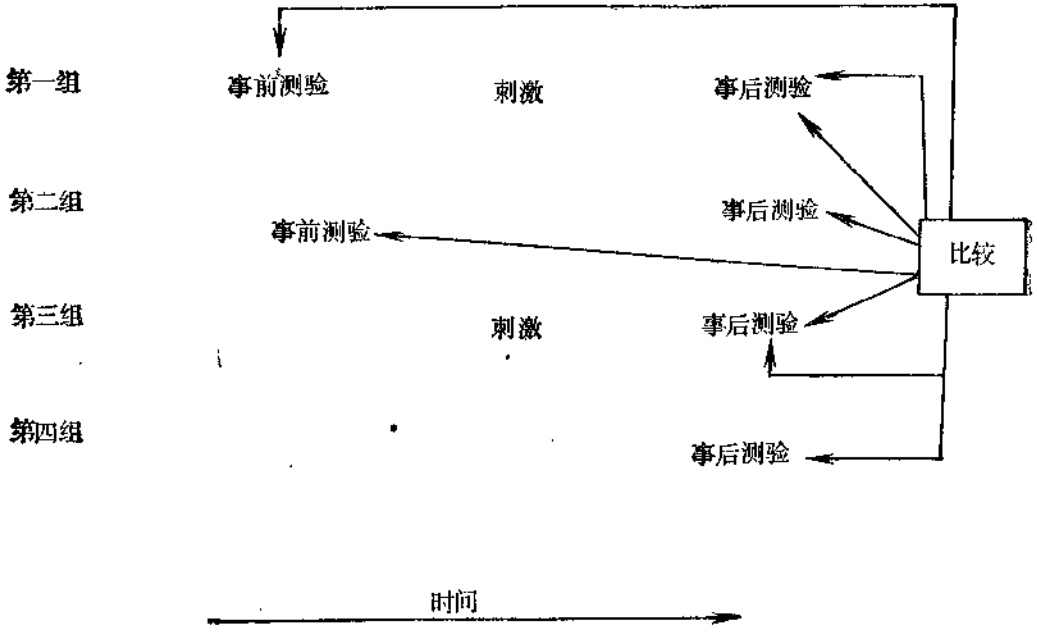


图8—5　所罗门四组设计

5 自然性质的实验

虽然人们常常把“实验”和“实验室实验”视为同义语，但许多重要的社会科学实验都是在实验室外进行的。这种“自然性质”的实验有时由研究者，有时则由社会的决策人物在现实生活中进行。

例如，一场台风袭击了某城，人们受到程度不等的损失。我们可以研究人们在遭受自然灾害情况下的行为模式。如，那些损失严重的人是否比损失较轻的人变得对未来的灾害有更高的警惕性？我们可以在台风之后访问该城市的居民，调查受灾前后损失轻重不同的人的警惕有何区别。在这一研究中，我们只能利用自然的事件来进行实验，而不能想象为做实验去安排一场台风。

6 实验方法之得失

设置控制组的实验方法的主要优点在于使实验刺激这一变量及其作用独立出来。这一点在基本实验模式中显得最清楚。一组实验对象在进行实验前

具有某种性质，对他们施加了实验刺激之后，其性质就改变了。由于这些研究对象并未受到其他刺激，我们可以做出结论：这种性质的改变当归因于实验刺激。

实验方法的另一优点是：由于单个实验往往规模有限，只需要少量的时间和金钱，这就有可能在不同的时间重复这一实验（当然情况并非总是如此，但重复一个实验比重复一个调查要容易一些。）同其他学科一样，研究结果的重复可以增强其可信性。

实验方法的主要弱点在于其人为性质。实验室中的社会过程往往不能代表现实世界中的社会过程。回到我们常用的例子：一部关于黑人历史的影片可能确实降低了实验组对象的种族偏见程度，但这并不意味着在一般的影剧院放映同一部影片能降低一般公众的种族偏见程度。当然，在自然性质的实验中，这种人为性质带来的问题比实验室实验要小得多。

我们讨论了几种内部失真与外部失真的来源。考虑到这些因素，我们可以设计出在逻辑上克服这些问题的研究方法。这种逻辑上的严密性是实验方法的最大优点，是其他研究方法难以做到的。

第九章
调 查 研 究

调查研究是社会研究中最常用的观察方法。第七章已经谈过的抽样方法是调查的基本步骤。本章将讨论问卷的设计以及填写问卷的两种基本方式：研究对象自填问卷及由访员提问记录答案。我们将讨论这两种方法并比较其优劣。本章还将用少量篇幅介绍第二手资料分析，即对他人收集的调查数据进行分析的方法。对于学生和缺少经费的人，这种方法特别适用。

1 适于采用调查研究方法的研究课题

在描述性、解释性和探索性的研究中都可以运用调查研究的方法。在调查研究中，分析单位多为个人。虽然也可运用其他类型的分析单位，如小组

等，但以单个人作为调查对象还是必不可少的。例如，在某项调查中，分析单位是离婚事件，但调查问卷还是要由离婚当事人来填写。

调查研究是收集第一手数据用以描述一个难以直接观察的大总体的最佳方法。精心设计的概率抽样可以提供一个能够反映总体的样本；精心设计的标准问卷可以从不同的研究对象那里获得具有同样形式的数据。

调查研究方法还特别适用于对一个大总体的态度与倾向的研究。民意测验是运用这一方法的例证，如盖洛普、哈里斯(Hairs)、罗波（Roper）、扬克洛维奇（Yankelovich）等人所作的民意测验。

2 问卷设计

社会研究中有多种观察方法要用到问卷，但其主要运用于调查研究这一方法中。

（1）一般问卷格式

问卷格式的重要性不下于问题本身的重要性。不适当的格式会造成答案的遗漏、混淆，甚至导致

研究对象拒绝回答问题的后果。

按照一般规则，问卷应当整齐清楚，每一问题后的空白要留得尽可能大些。缺乏经验的研究者总是担心问卷看上去太长，于是把几个问题并在一行写，或简化提问方法，以为问卷的页数总是越少越好。这种作法是错误的。将几个问题写在一行中会使人忽略后面的问题；简化的提问容易引起误解；人们也并不喜欢页数虽少但第一页费了他过多时间的问卷，反而喜欢虽然初看页数不少但轻而易举就填完了第一页的问卷。

（2）答案格式

多数问卷都是要求答问者从一组答案中选择一个答案。最好是用方括号或圆括号限定答案的空间，而不要留下一片空白让答问者随便乱划（如图9—1）。此外，也可要求答问者图出他要选择的答案（如图9—2）。

〔　　〕是	〔　　〕是
〔　　〕不是	〔　　〕不是
〔　　〕未详	〔　　〕未详

图9—1　答案格式

1. 是

②. 不是

3. 未详

图9—2 圈出答案

（3）后续性问题

问卷中有些问题可能只适用于某种类型的人而不适用于其他人，如关于节育方法的研究不会问男人是否吃过避孕药。

这种情况常常发生在研究者要提出一系列问题时。如首先想了解答问者是否属于某一组织，如果回答是肯定的再问他多长时间参加一次会议、是否拥有办公室等等。或者研究者想了解答问者是否听说过某一政治观点，然后了解那些听说过的人对这一观点的态度。

这种系列问题中的附带问题又称后续性问题：答问者是否应当回答后续性问题是由他对第一个问题（前驱问题）的回答决定的。这样设计问题是为了使人们免于回答那些他们不了解的问题。

后续性问题有几种格式，其中如图9—3所示的格式是最为清楚有效的一种。这种格式的两个关键因素是：第一，后续性问题要用框子与其他问题隔开；第二，要用箭头将后续性问题与前驱问题中的

适当答案联在一起。如图所示，只有对这一问题作出肯定回答的人才需要回答后续性问题，其他人则可跳过这些问题。

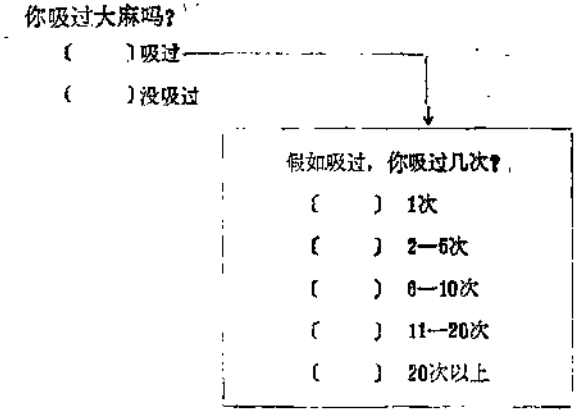


图9—3 后续性问题格式

对这一格式掌握纯熟之后，更加复杂的后续性问题也可设计得井井有条。图9—4就是一个更复杂的例子。

在后续性问题过长时可在问卷上注明哪些人可以跳过哪些问题不答，如图9—5所示。

（4）问题矩阵

当对几个问题的回答能够用相同的度量尺度时，可将问题设计为一个矩阵，如图9—6所示。

你听说过“模范城市试点”吗？

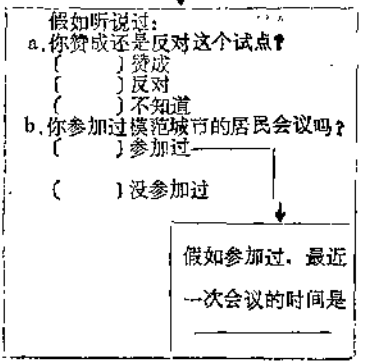


图9—4 复杂的后续性问题

你在选举中投票了吗？

( )投了（请回答问题14—25）

( )没投（请跳过问题14—25，直接回答问题28）

图9—5 关于跳过某些问题的说明

对下列各问题请表明你是强烈赞成、赞成、反对、强烈反对或未详

	强烈赞成	赞成	反对	强烈反对	未详
a.美国需要进一步加强法制	( )	( )	( )	( )	( )
b.美国警察不应当携带武器	( )	( )	( )	( )	( )
c.暴乱发生时，对哄抢者应当就地枪决	( )	( )	( )	( )	( )

图9—6 问题的矩阵格式

这种格式的优点在于：第一，节省空间；第二，答问者能较快填完这类问题。此外它还有利于同一矩阵中的几个问题进行比较。

使用这种格式的潜在危险是把应当做特殊回答的问题硬套在矩阵里，从而损害了答案的质量。此外，这种格式可能使某些答问者形成固定的答案模式，如：所有的问题都填"同意"。尤其是当几个问题很相像时，答问者以为所有的问题都差不多，因此一目十行，对问题做了错误的回答。解决这一问题的方法是变换问题的性质，避免将相似的问题放在一起，以及将问题设计得简短清楚。

（5）提问顺序

提问的顺序有时也会影响对问题的回答。如果一开始提出关于宗教信仰程度的概括性问题（如"一般地说，宗教信仰在你的心目中有多么重要？"）在回答后面关于宗教信仰的具体问题时，答问者就会尽量使自己的回答与前面的总评价取得一致。反之，一开始提出关于宗教信仰的若干具体问题，答问者也会努力使作为结论的总评价与前面的答案保持一致。为了克服上述两种倾向，有些研究者主张随机地安排问题顺序，结果并不理想，因为随机顺序的问题会使人感到混乱、无价值或难以回

答。

在讨论理想的提问顺序时首先应当对自填问卷与访问这两种方式加以区分。在自填问卷中，最好以最有趣的问题开头，使人看了头几个问题后愿意填写整个问卷。不要在一开始提出让人头疼或看了害怕的问题（例如提出关于性行为或吸毒的问题）；比较沉闷的一般性问题（年龄、性别等）应当放在结尾处。无经验的研究者往往把这类一般性问题放在开头，给人造成一种公文报表式的印象，因而对填写问卷失去兴趣。

访问调查则恰恰相反。当访员敲开一家的大门之后，必须很快赢得对方的好感。在简短地介绍了研究目的之后，他应当从家里有几口人，各自的性别年龄等等一般性问题开始提问。因为这类问题容易回答，对对方不造成任何威胁。一旦建立起友答的关系，访员才可以转向较为敏感的问题。一个劈头就问“你信不信上帝”的访员恐怕很快就会较灯发掉。

（6）答问指南

无论是自填问卷还是由访员填写的问卷都应当有清楚明白的答问指南。如“请在括号内划钩”等字句。如果问卷分为几部分，如政治态度、宗教信

仰、背景情况等，则应在各部分的开始介绍一下这一部分的内容与目的。尽管封闭型问题的各类答案是相互不包容的，有时答问者仍会感到合适的答案不只一个。如果研究者只要求选择其中一项，则应注明“请选择一项最佳答案”。如果要求答问者填写所有他们认为合适的答案，也应另做说明。

（7）问卷范例

图9—7是一份长达三十页的学生调查问卷的一部分，这个问卷是由夏威夷大学学生在1969年设计的。这项调查的目的是了解学生们在政治、宗教、教育等各个方面的态度和倾向。这一问卷并非十全十美，读者可以从中发现不足之处改进这一问卷使之臻于完美。

总说明：用铅笔或钢笔填写此问卷。多数问题只需要在适当的地方划“√”，只有少量问题需要写出答案。如果有附加的评论请写在页边。可忽略问题旁的数字，它们是为输入计算机设计的。

A.政治倾向

（1）请表明你对下列陈述是强烈赞成、赞成、反对、强烈反对还是不了解。

（6—15）

	强烈赞成 1	赞成 2	反对 3	强烈反对 4	未详 5
a.如联合国能成为世界政府是件好事。	( )	( )	( )	( )	( )
b.污辱美国国旗的人应当进监狱。	( )	( )	( )	( )	( )
C.美国军费过高。	( )	( )	( )	( )	( )
d.中央情报局权力过大。	( )	( )	( )	( )	( )
e.美国1945年对日本使用原子武器是正义的。	( )	( )	( )	( )	( )
f.我们应当支持国家政策，甚至当它是错误的时候也应当支持。	( )	( )	( )	( )	( )

（2）请表明你对下列越战政策的态度，赞成、不赞成或不了解。

（16—25）	赞成 1	不赞成 2	未详 3
a.保持美国军事行动的目前水平	( )	( )	( )
b.立即开始美军单方面撤军	( )	( )	( )
c.将美军撤往南越的战略村	( )	( )	( )
d.轰炸北越的战略目标	( )	( )	( )
e.美国地上部队进攻北越	( )	( )	( )
f.南越地上部队进攻北越	( )	( )	( )
g.对北越使用核武器	( )	( )	( )
h.在南越停止一切美国轰炸	( )	( )	( )
i.给美国军界领导人按其认为适当的方式进行战争的完全自由	( )	( )	( )
j.继续巴黎和谈	( )	( )	( )

（3）略　　（26—28）

（4）关于越战你还有什么看法？

（29—30）

（5）略　　（31—36）

（6）略　　（37）

（7）最近学校当局对是否应当允许征兵者和反征兵者进入夏威夷大学校园与学生交谈这一问题产生分歧意见。你个人认为学校当局应当做出下列哪种决定？

（38）

1 〔　〕应当允许征兵者和反征兵者二者与学生交谈
2 〔　〕应当只允许征兵者与学生交谈
3 〔　〕应当只允许反征兵者与学生交谈
4 〔　〕不应当允许征兵者和反征兵者与学生交谈
5 〔　〕我不知道。

（8）略　　（39—43）

（9）略　　（44—50）

（10）你认为哪个政府应当享有中国在联合国的代表权：台湾的国民党政府还是大陆的共产党政府？

（55）

1 〔　〕台湾国民党政府
2 〔　〕大陆共产党政府
3 〔　〕两个政府都应有代表权

*引者注：由于这一问卷多数答案空间在问题的右边，这两个问题也应保持一致，否则会给数据处理带来困难。

（11）请表明你对美国对华政策的态度

（56—61）

	同意 1	反对 2	未详 3
a.在外交上承认中国	( )	( )	( )
b.与中国进行贸易交往	( )	( )	( )
c.对中国进行经济援助	( )	( )	( )
d.与中国进行文化交流	( )	( )	( )
e.支持中国保存现有军事能力	( )	( )	( )
f.摧毁中国的军事力量	( )	( )	( )

（12）略（62—72）

（13）总的来说你如何评价自己与父母的政治倾向？

（73—75）

	自己	父亲	母亲
极右	1( )	1( )	1( )
非常保守	2( )	2( )	2( )
温和的保守派	3( )	3( )	3( )
温和的自由派	4( )	4( )	4( )
非常自由	5( )	5( )	5( )
极左	6( )	6( )	6( )
其他（请说明——）	7( )	7( )	7( )
未详	8( )	8( )	8( )

（14）你属于哪一党派？

（76）

1 ( ) 民主党
2 ( ) 共和党
3 ( ) 美国独立党
4 ( ) 和平自由党
5 ( ) 其他（请说明——）

6 〔 〕 无党派

（15）你在1968年11月时有选举权吗？

（77—78）

1 〔 〕 有选举权*
2 〔 〕 无选举权

*15a. 如果有选举权你投票了吗？

1 〔 〕 投了
2 〔 〕 没投

（16）无论你在1968年11月时是否有选举权你愿意选下列哪一位候选人为美国总统？

（79）

1 〔 〕 赫伯特·哈弗瑞
2 〔 〕 里查德·尼克松
3 〔 〕 乔治·沃伦斯
4 〔 〕 爱尔德里奇·克莱弗
5 〔 〕 无

（80/R）

图9—7 问卷范例

3 自填式问卷

自填式问卷除通过邮寄方法之外还有其他几种常用的方法：可以把问卷发给一组研究对象，在同一时间、地点收集填好的问卷；可以由研究者把问

卷送到研究对象家中，向其解释这项研究的目的，把问卷留下等过些时填好后再来取；也可以用邮寄的方法寄出问卷，然后由研究者去收回问卷。一般说来研究人员亲自去送交或收回问卷比单纯邮寄的方法回收率要高。

（1）通过邮政散发和回收问卷

通过邮寄搜集数据的基本方法是将一份问卷、一封解释研究目的的信和一个写好回信地址贴好邮票的信封同时寄出。人们往往不愿填写和寄回通过邮政散发的问卷，其主要原因是嫌麻烦。设想一位研究对象收到一份问卷，填好之后却需要自己找信封写地址贴邮票，他会愿意填写和寄回这样的问卷吗？因此应当尽量为研究对象的方便考虑。

（2）寄回问卷的启示

研究者应当认真记录问卷的回收率。一个很有价值的方法是制作回收率图。以问卷寄出的日期为第一天，记录每份问卷是在第几天寄回的。据此可制作两张图。一张是每天问卷寄回份数的分布图，另一张是问卷寄回份数的累积比率图。其作用是使研究者了解事情的进程，并为加寄催促信的适当时间提供依据。

无论散发问卷时是否已有号码，每当一份问卷寄回要给它一个新的登记号，登记号应是连续的。这样做有很多好处。例如，研究者要做一项人们关于某个政治人物的看法的调查。在收集数据过程中，报刊上忽然披露了这个政治家有情妇的秘密。如果研究者了解这一消息见报的日期和问卷寄回的日期，即可推测出这一消息对该政治家声望所产生的影响。

（3）催促信

催促的方法有多种，最简单的一种是寄一封催促信，敦促研究对象尽快填写问卷。更好的办法是随催促信一起再寄一份问卷，因为有些人之所以未填问卷是由于邮寄丢失或错误投递。

方法论文献表明，催促信是提高邮寄调查问卷回收率的有效作法。在实践中，邮寄三次（初次邮寄加两次催促）是回收率最高的作法。每次投邮的间隔以二到三周为宜。

（4）可接受的回收率

缺乏经验的研究者常常提出的一个问题是：邮寄式调查应当以多大的回收率为可接受的回收率。从理论上说，调查分析所使用的推论统计学是建立

在样本的所有成员都填写并寄回问卷这一假设之上的。由于这一点是绝对做不到的，就不得不考虑回收偏差的问题。研究者应当检验回收样本与初始设计的差距，看其是否仍然符合随机样本的要求。

回收率直接影响样本的代表性。高回收率造成的偏差显然低于低回收率造成的偏差。但回收率应当高到什么程度才是可接受的呢？一般认为百分之五十的回收率是可以用来分析和报告的起码比例；百分之六十的回收率是好的；百分之七十就非常好了。但这只是约定俗成的规定，并无统计学的依据。

4 访问调查

访问是收集调查数据的另一方法。研究者派出访员向研究对象口头提问并当场记录其答案。访问一般是面对面的，但也有通过电话访问的。访问一般需要一个以上或更多的访员。

（1）访员的作用

由访员填写问卷的方法比起自填问卷的方法有多种长处。首先，回收率高。一个正确设计和实施

的访问调查回收率应当至少达到百分之八十到八十五之间。答问者不好意思拒绝站在门口的来访者，却比较容易扔掉通过邮政寄来的问卷。

从问卷内容看，访员到场一般会减少填写“未详”或空着不填的项目的比例。访员访问这一方式还可减少答问者对问题的误解。如果答问者对问题发生误解或看不懂问题，访员可以加以解释，从而得到适当的回答。此外，在提问的同时，访员还可进行观察。例如在人种的问题十分敏感的情况下，访员可直接观察答问者的人种。其他如住房质量、答问者的英文口语能力及其对研究的一般反应等等也可进行直接观察。

访员的存在不应当影响答问者对问题的理解和回答。因此，访员应当成为问题与回答之间的中性媒介。如果成功地做到了这一点，不同访员对同一研究对象的访问应当得到完全相同的答案。这一中性原则在地域样本中尤其重要。为了节省时间金钱，某访员往往要负责某一地理区域上的全部访问，如一个街区或附近的几个街区。假如该访员用自己的观点影响了答问者，这种偏差可能会被误认为这一区域的特殊性质。

假定某项对建造廉价住房的态度的调查旨在为政府的建筑计划选择适当地点。负责某街区的访员

可能通过语言或手势表示出不喜欢这一计划，并影响了答问者，从而使调查结果显示出这一街区的居民强烈反对在该街区建造廉价住房，其实他们的反对态度可能仅仅反映了访员本人的态度，这就违反了这项调查的本意。

（2）访问的一般规则

虽然访问的方式应当依调查总体与调查内容的性质而定，但一般情况下的访问还是有规则可循的。

〔外貌与举止〕按照一般规则，访员应当穿得与被访问者相似。一个穿着过于华丽的访员很难得到穷人的合作；而一个穿着过于寒酸的访员会在富人那里遇到类似的困难。访员的衣着服饰应当尽量做到干净整洁。虽然中产阶级的干净整洁不一定能为社会的各个阶层所接受，但它代表了一般的道德规范并能被大多数人所接受。衣着服饰常常是人们态度和倾向的外部标志。穿花衣服、留长头发、留大鬓脚、留大胡子可能是政治左倾、性行为随便或吸毒等倾向的反映，虽然并非总是如此。

访员的举止应当令人感到愉快。由于访员要问及答问者的个人生活与态度，他们必须显得对答问者饶有兴趣但又不至于被疑为间谍。访员的态度应

当是放松的和友好的但又不能太随便或太固执。一个合格的访员应当有能力迅速确定答问者最容易同什么样的人相处，最喜欢同什么样的人交谈。显然，如果访员能成为答问者喜欢与之相处的人，他的访问会更加成功。

〔熟悉问卷〕如果访员不熟悉问卷，就会加重答问者的负担，访问会占用更多的时间而且很不愉快。因此，访员应当逐个认真看懂每个问题并练习大声朗读这些问题，最终达到可以流利无误地读出每个问题的程度。访员的最高境界是象演员在戏剧电影中念台词，应当象谈话一样自然，用的又完全是问卷上的词句。

访员还应当熟悉问卷上所有的说明。总会有一些问题与答问者的情况有差距，访员必须决定在这种情况下如何对问题作出解释。在问卷的总说明及特殊说明中应当有有关的解释，访员应当对之十分熟悉，以便在需要时随时引用。

〔使用原话〕第六章曾讨论过使用词语的重要性。词语上的轻微变化会导致答案大相径庭。无论研究者在问卷的词语设计上下了多少功夫，一个用自己的话来提问的访员会使他的一切努力化为乌有。

〔准确记录〕当问卷中有开放型问题时，访员

必须确切地记录答案，不要企图去总结、分段或改正语法。这一点极其重要，因为访员并不了解对这些答案将如何进一步处理。甚至连研究者本人也还没有决定进一步处理数据的方针。例如，问卷询问对交通状况的看法。某人回答说路上车辆太多，应当想办法限制车辆的数量；另一人则认为应该扩建道路。假如访员用同一个总结式语言“交通拥挤”来记录这两个人的回答，研究者就不能得到这两种答案最初所包含的信息了。

有时答问者可能极其不善于表达，语言含混不清。如果访员能从答问者的手势或声调上理解其意就应当在准确记录回答的词语部分之外，在页边写下自己的理解和作此种解释的原因。这些页边的评论是很有价值的，如答问者不知如何回答，表现出愤怒、窘迫等等。

〔追问答案〕有时答问者会答非所问。例如，问题是要求表明对某项说法是强烈赞成、赞成、反对、还是强烈反对，但答问者却只说了一句“我认为这种说法是正确的”。这时访员就应追问一句：“你是强烈赞成呢还是一般赞成？”如果必要，访问可以解释一下必须从几类答案中选择一种的道理。如答问者坚决拒绝选择，访员则应当准确记录答问者回答的原话。

开放型问题较多用到追问。例如对关于交通状况的问题，答问者可能仅仅说："很糟"。访员追问的方法是多种多样的。最好的追问方式是沉默。如果访员默然无语用笔作出准备记录的样子，答问者就很可能打破沉默作出进一步的评论（这个技巧是新闻记者最常使用的）。适当的追问句可以是"怎么？"或"怎么个糟法？"最常用的追问句是："还有呢？"

为达到研究目的常常需要追问，然而一个绝对必要的原则是，追问必须是中性的，不能影响问题的性质。如果研究者在设计问卷时能够预料到某问题需要追问，则应在问卷的问题后设计追问短语。这样做有两个好处：第一，由于时间充裕可以将追问句设计得更好；第二，由于所有的访员都使用相同的追问句，即使追问没有做到完全中性，所有的答问者受到的刺激却是相同的。当然问题设计应尽量避免偏颇，但更重要的是每个人都回答同样的问题，哪怕是有些偏颇的问题。

（3）协作与控制

多数访问式调查需要多个访员的帮助。在大规模的调查中，访员是被雇和有报酬的。在这种情况下基本功在于善于控制。控制包括两个方面：训练

访员和指导他们工作。

访员的训练应当是集中的而不是个别的训练。首先对他们说明这项研究的目的。尽管访员只参加收集数据阶段的工作，使其了解整个研究的目的是有益的。如果访员不了解自己是在干什么，就会情绪低落缺乏动力。访员训练从讨论前边讲过的一般规则开始，然后转向问卷本身。研究者要大声朗读每一问题，解释其意义，并回答访员提出的一切问题。

要预先想到一切可能遇到的问题，对如何处理这些问题做出特别说明，一个简单的例子是某人说他下周就满25岁，访问可能不知把他算作24岁还是25岁更合适。特别说明就要对此类问题作出规定（如规定一律以上一个生日的年龄为准）。应向访员讲解所有的特殊说明。

问卷讲解完毕，研究者应当为访员做一两个示范访问。示范越接近现实越好。不要在中途停下来解释自己为什么这样或那样做，事后再做解释。因为当答问者回答问题时，这种插入性的解释会使他觉得访问是个演习而已，因而影响访问效果。

做完示范后应将访员分为两人一组，一问一答。全部问卷问定后再让他们交换角色来问。研究者从旁观察。全组练习完毕应当让访员谈谈经验体会并

提出进一步的问题。

训练的最后阶段是模拟访问。随便找一些人来访问，但一定不要用样本中的人。每个访员做了三至五个访问后交回问卷，检查全部合格后即可交给他们真正的访问任务了。不要一下全部撒手不管，而应当要求访员每访问20至30人来报告一次。研究者要仔细检查问卷的填写情况以便及时发现问题。

（4）电话调查

电话调查的结果在理论上只能推论到有电话的人这一总体。数年前，由于穷人家没有电话，这种方法曾造成重大的社会阶级偏差。然而现今电话已成为绝大多数美国家庭的标准设备。据统计局估计，在1978年，97%的家庭拥有电话。因此，阶级偏差明显地减少了。

一个与抽样有关的问题是如何处理那些电话簿上查不到的号码。如果调查样本是根据电话簿抽取的，那些不愿公布自己电话号码的家庭——一般是富人——就会被漏掉。这个潜在的偏差已经从技术上解决了：运用随机号码这一最新的电话抽样方法，即根据电话号码的位数抽取一组随机数字。电话接通后与家中的哪个成员交谈也可根据随机做字决定。

电话调查比面对面调查要短，一般不超过10—15分钟。电话调查的优点在于省工省时。此外访员用不着顾虑衣着服饰。有时，答问者的回答甚至比面对面调查更诚实一些。

5 两种调查方法之比较

自填式问卷一般比访问调查更节省金钱和时间。对于缺少经费的调查者，如学生，这一点就很重要。而且自填式问卷适用于较为敏感的问题。在匿名的自填式问卷上答问者更容易说真话。

访问调查的优点在于问卷回收率高，减少对复杂问题的误解，以及在回答问题之外进行直接观察等等。研究者应当根据研究的需要和经费情况在这两种方法中取舍。

6 调查研究方法之得失

同社会科学研究的其他观察方式一样，调查研究也有其特殊的得与失。了解这些得失对于确定是否在某项研究中采用调查的方法是很重要的。

调查的方法特别适用于描述一个大总体的性质。一个认真抽取的概率样本加上一个标准化的问卷可以提供对某学生群体、某城市、某国家或其他大总体的一般描述。

调查的方法，特别是自填问卷式的调查，可以采用很大的样本。样本容量高达2,000人的调查是很常见的。而大样本对于描述性或解释性分析都是十分重要的。从某种意义上说，调查的方法较为灵活，可以对一个题目提出多种问题，做多种分析。此外，标准化的问卷便于度量和进行数据处理。

但是调查的方法也有不少缺点。第一，标准化的要求常常导致削足适履。在这一意义上，调查的方法常常使复杂的题目流于表面化。其次，调查很难触及社会生活的背景。它不能象参与观察的方法那样，提供对研究对象在其中思想和行动的生活环境的感觉。此外，调查象实验一样具有某种程度的人为性质。在问卷上做出保守回答的人不一定是保守派，给出种族偏见答案的人并不一定真有偏见。这一缺点在社会行为研究的领域尤其明显。调查不能度量社会行为，而只能收集对行为的自我报告。

调查方法的一般特征是准确性较低而可靠性较高。例如，与实地研究相比，调查的人为性质限制了它的准确性。但是由于对所有的调查对象施加标

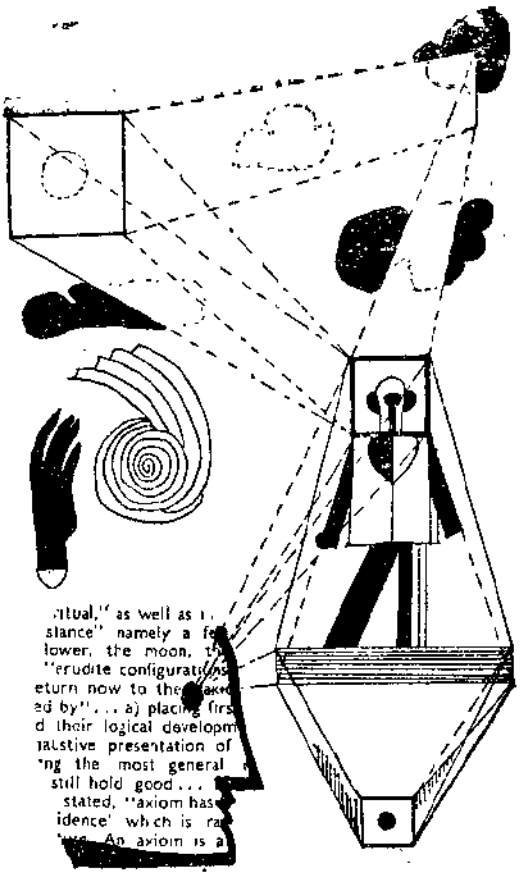
准化的刺激，调查可以避免不可靠的观察。

7 第二手资料的分析

作为一种观察方式，调查研究包括下列步骤：（1）设计问卷，（2）抽取样本，（3）通过访问或自填式问卷收集数据。一个大规模的调查往往要花数月甚至一年以上的时间。有一种方法可以避免这种时间和金钱的耗费，即第二手资料的分析。

由于计算机在社会研究中的应用，使社会研究者得以分享大量数据资料。从六十年代开始，第二手资料的研究在国际规模上得到发展。一些研究中心相互协作，形成了数据档案网。录好数据的磁带和输入卡被存入档案库，象传统的图书馆收藏图书一样，供广泛交流与使用。不同的是图书馆出借书籍，而数据资料则需要复制和购买。

第二手资料分析这一方法的优点是显而易见的，但它也有缺点，主要问题在于准确性。某个研究者为其特殊目的收集的数据不一定与另一研究者的研究兴趣相符。


ritual," as well as
slance" namely a
lower, the moon,
"erudite configurat
eturn now to the
ed by"... a) placing
d their logical develop
austive presentation of
ng the most general
still hold good...
stated, "axiom has
idence' which is
An axiom is

第十章
实地研究

实地研究包括下列研究方法：参与观察、直接观察和个案研究。实地研究方法在社会科学中是一个既老且新的方法。多少世纪以来，它一直为社会研究者所使用；但近年来，它在方法论中受到注意并得到重新定义。实地研究方法是所有的人无时无刻不在使用的方法。从某种意义上说，无论何时我们观察或参与社会行动并试图去理解它，就是在做实地研究了；无论何时我们向其他人讲述自己的观察，就是在报告实地研究的结果。

实地研究方法与其他形式的观察的区别在于它不仅仅是收集数据的活动，又常常是产生理论的活动。实地研究者很少是带着要加以检验的假设开始研究的，而是常常从正在进行的社会过程中发现有意义的事情，而这些事情是事先难以预知的。实地研究方法的基本作法是，首先做初始的观察，从中发

现出暂时的一般性结论，这些结论启示研究者应当做哪些进一步的观察，而进一步的观察又导致结论的修正。归纳与推演这两项原则的实践在实地研究中比在其他一切方法中都显得更明显和不可或缺。

1 适于采用实地研究方法的研究课题

实地研究的主要长处之一在于它可以使研究者看到事物的全貌。直接全面地观察社会现象可以使人对事物有深刻和充分的理解。这种观察方式适用于难以简单运用度量手段的课题。

与其他方法不同，实地研究方法特别适用于那些只有放在一定自然背景中才能很好理解的态度和行为。在某种人为的背景下用试验和调查的方法来度量行为和态度是可以的，但这种方法不是对一切行为都是最佳的。比如对某些宗教仪式的观察，实地研究的方法就更好。

此外，实地研究方法特别适用于研究随时间推移而变化的社会过程。实地研究者可以在暴乱发生时对它进行研究而不必事后再回过头来研究。

实地研究的方法还适用于研究学生示威运动、法庭程序、劳资谈判、公众听证会以及类似的在有

限的时空发生的事件。

在《社会环境研究》一书中，约翰·洛夫兰把适宜使用实地研究方法的社会现象归纳为以下六种：

（1）事件。转瞬即逝的、只持续几秒、几分钟几小时的事件。

（2）活动。持续时间较长——数天、数星期或数月——的行为。

（3）意义。参与者对其行为的定义和解释。

（4）参与。人们对某种状态或环境的卷入和适应。

（5）关系。人们的相互关系。

（6）环境。被研究事物的全部环境。

2 观察者的各种角色

瑞芒德·勾德讨论过实地研究者的四种不同角色：完全参与者，观察的参与者，参与的观察者和完全观察者。完全参与者可以是他所研究事件的真正参与者（如一个校园示威的参加者）或装作是一个真正的参与者。在这两种情况下，你象一个真正的参与者那样行动，别人只知道你是参与者而不知道你是研究者。

这里有一个伦理的问题：当你进行研究时，设法获取人们的信任而使他们告诉你一些绝不肯向身份明确的研究者吐露的事情，这种作法符合道德吗？没有任何一个研究者是为欺骗而欺骗的。他们这样做是基于下列信念：如果研究对象不知道他们被研究则会表现得更加自然、诚实。假如人们知道自己是研究对象，他们首先就有可能把研究者赶走；他们也可能改变自己的言行使之显得更值得尊敬；还可能使被研究的社会过程发生重大的改变。例如，一群学生本来打算烧掉学校行政大楼，但由于发现有人在研究他们因而放弃了这一行动。

这里还有一个科学性的问题。假如你是个完全参与者，就会对正在研究的事物发生影响。作为一个参与者将必须参与，而这种参与就可能影响研究对象。例如，别人问你下一步该如何行动，不管你怎么说都会影响事情的进程。假如全组听从了你的建议，你明显地影响了事情的进程；假如大家拒绝了你的建议，这一否决又可能对下一个决定产生影响；如果你不表态，则有可能在组内增加一种犹豫不决的气氛。这种影响简直是完全不可避免的。

出于这种伦理和科学的考虑，实地研究者常常放弃去扮演完全参与者的作法而选择另一种角色：观察的参与者，参与全部活动同时说明自己在做研

究。然而，这种角色也有它的问题。别人可能把注意力从自然的社会过程转移到你所作的研究上去，结果观察到的事物就不那么典型了。此外，观察的参与者很可能会“入乡随俗”，从而丧失了科学的超然态度。

参与的观察者以研究者的身分出现，与社会过程的参与者交往，而并不装扮为参与者。这种角色的范例是新闻记者。

完全观察者对社会过程做纯粹观察而并不打算成为这个过程的一部分。由于这种观察者尽量不引人注目，研究对象常常并不知道自己在被人研究。例如，观察者坐在一个汽车站旁边观察不按交通规则过街的行为。完全观察者既不会影响观察对象也不会象完全参与者那样陷入被观察的事物之中，但缺点是难以得到对研究对象的充分了解因而所作的观察较为粗浅。

不同的情况下需要采取不同的角色，可惜并没有关于如何作选择的指南。研究者只能根据自己对情况的了解和判断做出选择。

3 实地研究的准备工作

假定你决定对某校的政治组织作实地研究。你既不是这个组织的成员也不大了解它的情况。你准备在一开始就向那个组织说明自己的研究者身份。

实地研究的准备工作首先是阅读文献。查阅所有关于这个组织的资料可以为你修定研究的理论框架提供依据，以便确定最值得研究的问题。

查阅文献包括两个阶段：（1）找到适当的起点，（2）“滚雪球”式的扩大检索。具体作法是，首先，通过查阅书籍和期刊索引找到一批书籍和文章，然后用两种方法扩大线索：第一，从书籍文章后面的文献目录上找资料索引；第二，在找到某一书籍期刊后，浏览左近的书籍，因为图书馆都是按类别存放书籍的。这就是所谓“滚雪球”了。

研究准备工作的下一个阶段是利用信息提供者。你可以向已经做过同类研究的人请教，或找那个政治组织的成员谈一谈。在正式与这个组织接触之前，你对它的一般背景应当是相当熟悉的。

接触研究对象的方法是多种多样的，按你的准备以什么面目出现而定。如果你想作为一个完全参

与者去研究饭馆刷碗工，最直接的办法是去找个刷碗的工作。接触某个学生政治团体的最简单方法是参加这个组织。对社会过程的实地观察也是如此。例如想观察示威游行就到示威地点去，想观察违章过街者就到行人拥挤的街道去进行观察。

在接触研究对象时，应当与他们建立友好的关系。如请组织的某个成员把你介绍给这个组织，向他们解释你的研究目的等等。解释到什么程度为宜则依研究的性质而定。

4 实地研究中的抽样

前面论述过的抽样技巧一般不适用于实地研究，因为实地研究中的总体和分析单位不太明显。假如你研究某城市的三个青少年帮伙，分析单位帮伙、单个成员还是城市？你只想描述这个帮伙还是想研究这类青少年帮伙的一般关系？重要的事情在于你的研究结果要推论到什么样的总体，而这个问题的答案在实地研究中往往随研究过程而改变。

实地研究观察对象的一切情况，在这个意义上它不算抽样，但事实上研究者不可能观察到一切现象，在这个意义上，他所观察到的部分事实上又是

从所有可能的观察中抽出的一个样本。例如观察者在一个动乱中观察暴力行为，观察到的行为则是一切此类暴力行为的样本。这个样本虽然不是事先设计好的，但还是应当服从于具有代表性这一原则。

麦考和塞蒙提出三种适用于实地研究的抽样方法：定额抽样、雪球抽样和特异个案。

首先，如果被研究的组织和社会过程的成员已有明确的合理分类，可以运用定额抽样的方法：选择不同类型的成员。如研究一个组织，可以分别访问领导人和一般成员；研究学生政治团体可以分别访问激进派成员和稳健派成员；或分别访问男人和女人，年轻人和老年人等等。

其次，雪球抽样。假如你想了解一个宗教团体如何招收新教徒，可以向近期加入教派的信徒询问他们介绍人的情况，然后再去追寻介绍人的介绍人，于是样本就象雪球一样越滚越大。

此外，麦考和塞蒙强调了特异个案的重要性。对于脱离正常模式的个案的研究可以加深对人们态度及行为的正常模式的理解。例如你想了解某学校在一个誓师会上表现出来的校风，去访问不参加誓师会的学生可以得到重要的信息。

5 提出问题

从某种意义上说，实地研究就是去看和听，但有时也有必要提问题和作记录。与调查方法中的事先设计问卷不同，实地研究一般采用不作事先设计的访问。

非事先设计的访问是访问者与研究对象之间的交流。在访问过程中，访问者只有一个大致的调查计划而并没有事先准备好的问题。访问者掌握谈话的大方向，主要听研究对象讲。

实地研究的特殊优点在于它的灵活性。对第一个问题的回答可能引出第二个问题。不应当仅问事先准备好的问题而应随问随听随想，不断修正和加深问题的提法。不要轻易打断研究对象的话，但是要注意因势利导。

6 记录观察内容

实地研究的基本工具是笔记本——或实地观察日志——和一枝笔。录音机和照相机有时反而不能

捕捉社会过程的一切方面。如果可能应当在观察时当场记录，如果做不到则做事后记录。

笔记应当既包括观察又包括理解。应当记录你“知道”的，同时记录你“想到”的。例如你记录某人发表的反对领导的意见而你认为他是想争夺领导权。应当把这个看法也记下来。

有时，准备一个标准的记录表可以简化记录工作。例如准备一个按年龄、性别、民族等分类的记录表。另外可以准备一些速记符号以便加快记录速度。

做好记录需要一些特殊的技巧。第一，不要过于相信自己的记忆力。即使你是过目不忘的人，也是以当场记录或事后尽快记录为好。第二，分段记录。一开始做一个粗略的记录（词和短语），然后重新整理一遍。第三，关于记录的详简：值得把观察到的一切细节都记录下来吗？回答是：值得。因为在实地观察时你很难确知事情的重要程度。一些初看不重要的细节后来可能被认为是很有价值的。

7 数据处理

虽然某些定量的分析方法可以用于实地研究，

但实地研究一般被视为定性的社会科学研究方法。如何处理定性的实地研究数据呢？

（1）整理笔记

观察后尽快整理笔记对于一个实地研究者是绝对重要的。应当以观察记录为线索，回忆整理出详尽的笔记，而且应当搞成一式两份以便写研究报告时剪贴用。

（2）建立档案

应当保留一份标明日期顺序的笔记，以后的档案都要建立在这个原始记录的基础上。到此为止我们一直在讨论观察，但包罗万象的原始观察并不能增加我们对社会生活的一般价值的理解。必须对这些原始观察进行分析和解释，从中发现行为的模式和意义。建立档案就是这种发现过程的第一步。

档案种类很多，建立何种档案应依研究性质及数据分析的需要而定。第一类是背景档案。假定你想研究一个社会运动，为其历史建立单独的档案是很有必要的。其中包括这个运动于何时何地发起，参加人数，运动史中的主要事件及其发生时间等等。

第二类是人物档案。其中包括运动的主要人物

的材料，甚至可以给每个重要人物单独立案。

第三类是文献档案。其中应当包括研究过程中所使用的一切资料的目录，免得写到研究报告的参考书目部分时需要到图书馆重新核对书目。

第四类是分析档案。如你认为这个政治运动具有宗教意义，于是就可以建立这个方面的档案，档案题目定为“宗教意义”。原始笔记中一切与宗教意义有关的内容——如一个研究对象认为这个运动提供了新的生活意义——都集中起来存入这个档案。又如你对运动使用暴力的程度感兴趣，可以建立以“暴力程度”为题目的档案，把笔记中一切有关暴力的内容存入这个档案。类似的分析档案还可建立许多。

8 数据分析

在实地研究中，观察和分析是紧密结合在一起的，按照一般规则，你希望在观察的现象中发现相同点和差异点。一方面，你在被观察的事件中寻找共同的模式，用社会学的术语则称为行为的规范，即研究对象的共同行为模式。在政治集会中是否所有的参加者都采取相同的行动对讲演人表示支持？

在宗教福音布道会上是否所有的参加者都在适当的时候说“阿门”？是否所有的妓女都爱穿戴富于诱惑性的服饰？实地研究者试图从中发现具有普遍性的现象。

另一方面，发现差异，即违反一般规范的异常行为。在福音布道会上，多数人齐声说“阿门”，但有少数人没有说。他们为什么违反行为规范？他们与其他人还有哪些区别？

有时，很难从某种行为中找出共同规范。如一群人在排队买电影票，有的呆立在那里，有的开始和陌生人攀谈，有的自言自语，有的不断探头张望队伍是不是往前挪了，有的不断地数手里的钱，还有的在看书等等。在这种情况下，重要的是为这些行为分类，然后尝试去发现与各类行为有关的特点：是富人还是穷人更爱在排队时数钱？是男人还是女人更容易和陌生人交谈？是老年人还是年轻人更爱自言自语？总之，你的目的是发现一般的行为模式。

对于实地研究者来说，理论命题的形成，实际事件的观察及其结论往往在同一过程之中。一般说来，不是用观察来检验某个理论，而是在观察的过程中建立理论。从观察中得出暂时的结论，这个理论框架又转而指导进一步的观察。这种收集数据与分析数据之间的相互作用，使实地研究变得比其他

研究方法都更灵活。研究设计因而可以随观察的过程不断加以改进。

9 得出结论

虽然没人提出过分析定性数据的全套逻辑规则，但豪渥德·卡海恩在《逻辑与现代修辞学》一书中对人们易犯错误的概括是很有价值的。他们提出的问题有：

〔偏狭〕人们往往透过某种历史和现实的有色眼镜来看世界。实地研究者也很可能会从自己认为有意义的方面来解释人们的行为。例如，基督徒从基督教的意义上看事物，社会主义者则从社会主义的角度看事物。

〔分类误差〕仅仅通过对观察的分类往往就可得出结论，而定性的量度很容易自觉不自觉地产生偏差，因此应当仔细检查度量尺度的正确性。

〔原因误差〕当你认为x是y之所以产生的原因时，应当自问情况是否当真如此，还有哪些其他原因？例如，当一个公司破产时，人们往往归咎于经理缺少经营管理能力，而不去考虑破产发生在严重经济衰退过程中这一因素。

〔忽略潜在因素〕实地研究者通过直接观察访问和在图书馆查阅资料，积累了大量信息。在作结论时需要对这些信息进行筛选。关于哪些信息在衡法之列显然要做权衡。例如，研究者作出结论：某新纳粹组织的成员之所以歧视黑人是由于害怕他们成为经济上的竞争者，于是我们推测这个组织的成员应大多属于工人阶级或中产阶级的下层。如果研究者忽略了关于这些人社会地位的资料，就很难使人相信前述结论。

〔虚假对立〕某些研究结论常常强调某一观点而排斥其他一切观点，例如“男性占统治地位的现象应当由经济学而不是生物学来加以解释”这一说法就把事情归结为仅仅有两种可能性：或是由生物学或是由经济学来解释，而认为从经济学角度才能真正解释男性占统治地位这一社会现象。然而实际情况是，除这两种可能性之外还有其他许多可能性，如社会传统、宗教信仰等因素都可能对男性统治作出解释。前述论断忽略了其他可能性而犯了“虚假对立”的错误。

10 实地研究范例

社会学研究生阿尔弗莱德（Randy Alfred）在1967年接受了研究和报告撒旦教会内情的任务。这个公然宣称与撒旦结盟反对基督的教会在旧金山设立总部，其领导人安东·莱维（Anton Lavey）具有超凡魅力。阿尔弗莱德于1968年4月开始接近这个教会并很快表示愿意入会。他的假话被信以为真，他不但入了会而且很快被提拔到领导岗位。他参加了许多教会的集会和仪式。到1973年，他共用了六百多个小时进行观察，同时阅读了关于这个教会的大量文献以及教会的出版物。他发现人们参加教会有多种多样的动机，有的为了纵欲，有的希望掌握魔力，有的是为了反抗传统宗教或一般社会传统，更有些人以为撒旦崇拜是未来世界的新潮流。在研究接近尾声时，阿尔弗莱德越来越后悔当初隐瞒了自己的真实身份，认为自己的作法不道德，因为人们信任了他并给了他负责职务。最后他去见莱维向他讲明了实情并请求允许发表自己的研究报告。莱维说，他早就怀疑阿尔弗莱德在搞研究。这样做是否不道德呢？不，莱维说，这正是撒

旦式的行为。

11 实地研究方法之得失

如前所述，实地研究特别适用于研究人们言行的微妙差别和随时间发展的社会过程。这种方法的长处在于可以使研究具有相当的深度，而许多其他研究方法都容易流于表面化。灵活性是实地研究的另一长处，研究设计可以随时改变。此外，实地研究可以节省经费。其他研究方法可能需要贵重的仪器和较高的人工费，而实地研究通常只有研究者本人和他的纸笔就够了。

实地研究也有一些弱点。首先，作为定性而非定量的方法，它很难产生对总体的精确描述。其次，从实地研究中得出的结论常常只是可能的而不是确实的结论。这一点在下面关于准确性、可靠性及概括性的章节还要详加论述。

准确性是指度量手段是否被用来度量研究对象而不是其他的事物；可靠性是指可靠程度——度量多次，是否每次都产生相同的结果；概括性是指特殊的研究结果是否适用于那些没有受到观察的人群和事物。

（1）准确性

调查和实验的方法常常容易流于表面而缺乏准确性。与之相比，实地研究似乎可以提供更加准确的度量手段。

“到现场去”是认识事物性质的有效方法。观察与概念抽象的过程本身往往就很有价值。此外，它们还可以为进一步的研究——不论是定性的还是定量的——打下基础。

实地研究便于使用综合的度量尺度从而使概念具有深度。例如，保守派和自由派的概念在调查和实验中是很难量度的，而在实地研究中，研究者可以不必去定义概念，而是给出详细的描述。

（2）可靠性

然而，实地研究有一个潜在的问题：可靠性问题。你根据你所了解的一切细节对一个好朋友的政治倾向所作的评价当然绝对不会是表面的，而是相当准确的。但难以肯定其他人做了同一观察后也得出同样的结论。

实地研究是深刻的，但同时又往往是带有浓厚个人特色的。每个观察者都有自己独特的立场，也许你认为你的朋友不过是个中间派，而我却认为他

是一个激烈的激进派分子。相形之下，比较的研究更可信些。例如说某人比某人保守。尽管你我政治倾向不同，在给一群人按保守程度排队时我们倒很有可能取得一致意见。

科学的一个主要目的在于对事物作出抽象概括。社会科学家研究那些特殊的情况和事件是为了对社会生活做出一般的理解。没有人真想了解某个个别的研究对象的情况。例如，没有人会去关心盖洛普样本中的1500个选举人中每个人各自想投谁的票，而仅仅关心是否他们的投票情况可以代表整个选举的倾向。

概括性是实地研究的另一个潜在问题。首先，如前所述，研究和度量的个人特色会影响研究结果，因而不同的研究者会得出不同的结论。假如观察结果因各个观察者而不同，那么它们与其被当作证据或事实还不如被当作某种看法的来源更为恰当。

其次，由于实地研究者对对象了解得既充分且深刻，他可以对对象做出非同一般的全面理解。然而正因为这种理解的无所不包的性质，它反而比一般正确抽样和标准化度量较少概括能力。

此外，在某项观察的过程本身也有一个概括性的问题。当你与某个研究对象交谈时，你并不知道

他的情况是否“典型”、是否具有对其研究对象的代表性。抽样偏差的潜在危险总是存在的。

如前所述，实地研究是社会科学家手中的一个有力武器，它与实验和调查的方法相辅相成。然而观察形式还不只这三种，在第三部分的各章将介绍其他的研究方法。

第十一章
非接触性研究

前面讲过的几种研究方法都要不同程度地接触和影响研究对象，实验居首，调查次之，即使是实地研究者也会在研究的过程中使研究对象发生某种程度的变化。但有些研究方法可以避免这个问题。

本章将讨论三种这类的研究方法：内容分析、现存统计资料分析和历史-比较分析。在内容分析中，研究者主要使用书面文献。例如你想比较美国三十年代和八十年代对国内国际问题强调程度的不同，就可以采用对这两个时期的报刊文章进行分析的方法。这一方法就叫作内容分析，即对信息的分析。

迪尔凯姆（Durkheim）研究自杀问题的方法则属于现存统计资料分析的方法。你会发现周围有大量的统计数据供你去理解社会生活。

最后，我们将讨论历史-比较分析的方法，在社会科学研究中，这种方法很早就开始使用，至今

仍在使用。象实地研究一样，历史-比较分析也属于定性的研究方法。运用这种方法进行观察和分析的主要资料来源是历史记录。用“比较”一词是因为社会科学家与历史学家不同，后者只是简单地描述历史事件，而前者则试图从这些事件中发现在不同时空条件下反复出现的共同模式。

1 适于采用内容分析方法的研究课题

内容分析的方法可以用在对任何信息形式的研究中，如书籍、杂志、诗歌、报刊、歌曲、绘画、演说、信件、法律等等。是否法国畅销小说中的爱情内容比美国的多？是否六十年代美国流行音乐的政治讽刺意味比德国的强？是否那些提出“面包与黄油”问题的政治竞选人比大讲最高原则的更容易获胜？上述每个问题都提出了一个社会科学研究的课题：第一个与民族性格有关；第二个与政治过程倾向有关；第三个与政治过程有关。对这些课题虽然也可以进行以人为对象的研究，但内容分析提供了另一种选择。

内容分析特别适用于对各类信息的研究并回答那个信息研究中的古典问题：谁说了什么？对谁？为什么？怎样？结果如何？作为一种研究方法，内

容分析要求解答什么、为什么及产生了何种结果这一系列的问题。

2 内容分析中的抽样

象以人为对象的研究一样，以信息为对象的研究也往往难以直接研究全部对象，因此需要抽样。

（1）分析单位

第四章曾谈到分析单位即在描述或解释研究对象时所使用的单位。例如，在计算家庭平均收入时，家庭是分析单位。我们询问每个家庭成员的收入情况于是个人成为观察单位，而家庭仍是分析单位。在比较各城市的犯罪率时，城市是分析单位。

在内容分析中，分析单位的问题显得比采用其他方法时更为突出。假定我们想了解法律量刑是否男女有别。在这一研究中，单个的法律既是观察单位又是分析单位。我们可以在多种法律中抽样，把它们按是否男女有别分类。假如我们想了解那些在法律上种族有别的州是否比其他州更有可能制定男女有别的法律，虽然在这个研究过程中还是需要研究单个法律，但分析单位却不是法律而是州了。

这个问题之所以重要是因为抽样取决于分析单位。假如分析单位是作者，就应当在作者中抽样；分析单位是书籍，就应当在书籍中抽样。但这并不是说抽样仅仅与分析单位有关。除了总体抽样外还常常需要做分类抽样。因此，假如作者是分析单位，我们可能需要：（1）从作者总体中抽取作者的样本；（2）从入选作者所写的书中抽取一个书籍的样本；（3）从每本入选的书籍中抽取一些章节用来作观察和记录。

（2）抽样技术

在对文献做内容分析时，抽样可以在下列各层次上进行：词汇、短语、句子、段落、章节、书籍、作者等等。对文字之外的其他信息形式也可以在任何概念层次上抽样。

第七章论述过的常用抽样技术在内容分析中全都适用。我们可以对法国或美国的小说家、密西西比州的法律、或莎士比亚剧中的台词做随机或系统抽样；我们可以把托尔斯泰的《战争与和平》每章的第23段抽出组成一个样本（注意运用随机起始）；我们还可以给披头士录制的每张唱片编号然后抽取一个容量为25的随机样本。

分层抽样也同样适用于内容分析。例如，要分

析美国报刊的政策评论，可先把报刊按地区、出版单位的规模、出刊频率等分层，然后从中抽取一个分层样本。其后还可以进一步对每份入选的报刊按日期分层抽样。

类集抽样也可用于内容分析。以前面举的关于报刊评论的研究为例，假如单个的评论是分析单位，那么在第一阶段对报刊的抽样就可以成为若干评论的类集抽样；在对政治演说的分析研究中，我们可以从对政治家的抽样开始，每个政治家就是若干政治演说的类集。如果小说是分析单位，我们可以从小说家中开始抽样，然后从每个入选小说家所写的小说中抽样，最后从每本入选小说的段落中抽样。通过对这些段落内容的分析对小说作出描述。

3 内容分析中的编录方法

内容分析的基本作法是编录，即根据特定的概念框架对信息——无论是口头的、文字的、还是其他形式的——做分类记录。例如，报刊评论可能被记录为自由的或保守的；小说可能被记录为宣传性的或非宣传性的，照此类推。

（1）明显内容和潜在内容

编录信息中的明显内容与记录标准化问卷的方法十分相似。例如要确定某些小说的色情程度，你可以直截了当地数一数“爱情”一词在每本小说中出现的次数，或平均每页出现“爱情”一词的次数。也可以用“爱”、“吻”、“拥抱”和“抚摸”这些词分别作为小说色情程度的指标。这种方法的长处在于容易和精确，使人确切知道你是在什么意义上使用“色情”这个词的。它的缺点在于准确性没有保障。“色情小说”这一用语所表达的意思要远远超过“爱情”一词出现的次数。

另一种作法是去记录信息的潜在内容：它的实际意义。如前例，你可以把小说从头至尾读一遍，或对段落章节进行抽样阅读，然后对读过的东西做出评价。显然，这种方法可以更好地理解信息的实际内容，但代价是不够精确。特别是当几个人对同一本小说进行编录时，他们可能会使用不同的定义和标准。同一段话，这个人认为是色情的，那个人认为不是。即使你一个人编录也很难数到定义和标准始终一致。此外，研究报告的读者也很难确知你所使用的定义。

解决这个两难命题的办法是兼用两种方法。假

如两种方法结果相似，这个结果就可以被接受；如果结果不同则应重新考虑整个理论框架。

（2）概念化和制定编录类别

任何变量的实施定义都是由多种属性组成的。这些属性应当是相互不包容和穷尽的。例如某篇评论不能既是自由的、又是保守的，虽然可以设置“中立”这一类别。也许对你的研究来说把小说分为色情的非色情的就够了，但也可以增加“反色情”这一类别。

内容分析可以使用多种度量层次。例如为报刊评论分类可以使用自由或保守这种定名性类别；或把它们从极端自由到极端保守做定序性排列。

对编录框架必须进行事前检验，即先对少数观察对象使用这一框架。假如你发现所有的报刊评论都被记入“自由的”这一类别，就需要重新考虑对这个属性的定义是否恰当了。

（3）计算和保存记录

如果你想对内容分析数据进行定量的分析，就必须使编录结果适应数据处理的需要。

首先，编录的最后结果必须是数字形式的。如果采用计算某些词句出现次数的方法，其结果当然

是数字形式的；对潜在内容的编录也应转化为数字形式，如1＝极端自由派，2＝温和自由派，3＝温和保守派等等。

第二，必须将分析单位和观察单位加以区分，特别是当二者不一致时。原始记录当然与观察单位相联。例如小说家是分析单位，小说是观察单位，通过对小说内容的观察对小说家作出分析。

第三，计算时绝不能忽略基数。如果只知道某个画家画了若干现实主义的画而不知道他一共作过多少画，就什么问题也说明不了。只有了解到在他

表11—1　　记录表（部分）

报刊编码	评论篇数	评价 1.极端自由 2.温和自由 3.中立 4.温和保守 5.极端保守	反共评论篇数	“赞成联合国”评论篇数	“反对联合国”评论篇数
001	37	2	0	8	0
002	26	5	18	0	6
003	44	4	2	1	2
004	22	3	1	2	3
005	80	1	0	6	0

的全部作品中一个很大的比例属于现实主义的作品，才能作出他基本上是一个现实主义画家的结论。同样道理，如果只知道“爱情”一词在某小说中出现了87次而不知道全书字数，则信息量几乎为零。

表11为报刊评论研究的记录表（部分）。

4 内容分析范例

内容分析的一个范例是对美国六十年代所面临的重大问题的研究。六十年代对于美国来说是一个骚乱的年代。这十年从关于公民权利和越战的辩论开始，以生态运动和女权运动的兴起而告终。

瑞·方克豪瑟（G·Ray Funkhouser)的内容分析是利用大众传播媒介（主要是周刊）来进行的。研究开始时，他所面临的问题有①研究所涉及的时间，②所用杂志，③抽样方法，④研究内容，⑤编录方法。他把时间定为1960年到1970年，选择了三种最著名的周刊；抽样方法简单之至：他使用了这个时期内出版的全部刊物，共1716期；他并未创造自己的编录类别，而是直接采用了《读者指南》的编目；他的作法是从那三种周刊中把每个条

目下的文章的篇数记录下来。表11—2是他的统计结果与盖洛普调查结果的对照。

表11—2 新闻杂志对六十年代各类问题的报道及盖洛普调查关于各类问题重要性的排列顺序

问题类别	文章篇数	排列顺序	盖洛普调查关于各类问题重要性之排列顺序
越南战争	861	1	1
种族关系（及城市骚乱）	687	2	2
校园动乱	267	3	4
通货膨胀	234	4	5
电视及传播媒介	218	5	12*
犯　罪	203	6	3
吸　毒	175	7	9
环境与污染	109	8	6
吸　烟	99	9	12*
贫　困	74	10	7
性（道德堕落）	62	11	8
女　权	47	12	12*
科学与社会	37	13	12*
人　口	36	14	12*

注：新闻报道的排列顺序与盖洛普调查关于各类问题重要性的排列顺序的相关系数为0.78（P=0.001）

* 这些项目在盖洛普调查中未列入“重大问题”故记为相同的顺序数。

5 内容分析方法之得失

内容分析的最大优点在于节省时间金钱，不需要大量研究人员和特别的设备。另一优点是保险系数大。假如在调查或实验中结果不理想，重做要耗费双倍的时间金钱；在实地研究中重做也许根本不可能，因为你研究的事件已不再存在了。而在内容分析中，弥补过失是很容易的。还有，内容分析可以研究在长时期内发生的过程。此外，它可以不打扰研究对象，不对它们发生影响，这是内容分析特有的优点。

内容分析也有缺点。它只限于记录下来的信息。准确性方面也存在问题，除非是研究信息交流过程本身。当然，可靠性还是很大的，你可以一再重复所做过的编录。这在实地研究中是不可能办到的。

6 现存统计资料分析

运用官方或半官方的统计资料进行社会科学研

究常常是可能的甚至是必要的。现存统计资料至少应当被当作数据的补充来源，它往往可以为研究提供历史背景材料，它也可以成为社会科学研究数据的主要来源。

（1）关于自杀的研究

人为什么会自杀？毫无疑问，每个自杀案件都有其独特的原因，但也可以按照某些共同原因为这些案件分类：经济上的失败、失恋、受辱、或其他个人问题。法国社会学家艾米利·迪尔凯姆(Emile Durkheim）曾经研究过环境特别是社会条件对自杀的影响。

在查阅了大量统计资料之后，各种模式逐渐显现出来，这些模式使迪尔凯姆大感兴趣。首先引起他注意的是自杀率的相对稳定。研究了几个国家，他发现自杀率年复一年总是几乎相同。他还注意到发生在炎夏季节的自杀比其他季节多得多。温度会不会对自杀有影响呢？如果事情果真如此，自杀率在南欧国家就应当比北欧国家高。然而他发现自杀率最高的国家是中等纬度的国家，因此温度似乎不是自杀的原因。他又探索了年龄的作用（35岁是最常见的自杀岁数）；性别的作用（男女自杀比例为4:1)及其他因素。最后他发现了一个重要模式：政

治动乱时期自杀率会突然增高。如1848年前后的欧洲国家。这个发现导致了一个假设：社会平衡的破坏会使自杀率增高；反之社会稳定与和谐似可防止自杀。

他还发现一个引人注目的模式：新教国家比天主教国家的自杀率高。在新教占统治地位的国家中，每百万人有190起自杀；新教-天主教混合国家是96起，而天主教国家只有58起。他注意到某些其他因素，如国家的经济文化发展水平也可能是这种区别的原因。假如宗教确实对自杀有影响，在各个国家内部，不同的宗教信仰也应带来自杀率的区别。经过研究，情况确实如此。于是他对宗教信仰对自杀有重要影响这一论断增强了信心。

回到更加高度概括的理论层次上，迪尔凯姆提出自杀是丧失规范和社会不稳定不和谐的产物。在政治动乱时期，人们会由于社会旧有方式的崩溃而感到迷惑和抑郁，从而走上绝路。相反，社会和谐稳定能防止自杀。同样道理，天主教作为一个结构远为严密一致的宗教体系比起结构松散的新教更能为人们造成亲密、和谐与稳定的感觉，因此，天主教徒的自杀率比新教徒低。由此，迪尔凯姆为社会科学的辞典增加了“社会反常状态”这个新概念。

（2）分析单位

现存统计资料分析的分析单位往往不是个人，正如我们在关于自杀的研究中所看到的那样。迪尔凯姆运用了政治-地理单位：国家、地区、州和城市等。对犯罪率、交通事故率等问题的研究也是如此。多数现存统计资料是描述群体的。

现存统计资料的这一性质带来一个问题：如迪尔凯姆想确定是否新教徒比天主教徒更容易自杀。资料并没有记载自杀人的宗教信仰。因而他不能肯定地说新教徒比天主教徒更容易自杀，而只能从新教国家、地区比天主教国家、地区的自杀率高这一现象进行推论。这个推论存在着一个潜在的危险，即群体层次上的行为模式可能并不反映个人层次上的行为模式。例如，在新教地区可能不是新教徒而是天主教徒自杀的更多，原因是那里天主教徒受到严重迫害而导致绝望和自杀。于是产生了这样一种可能性：虽然新教国家有较高的自杀率但并不是由于新教徒自杀造成的。

迪尔凯姆用两个方法解决了这个问题：第一，他的结论不仅建立在事实整础上而且建立在严密的逻辑基础上。理论和事实的一致使相反的解释难以成立。第二，通过不同途径深入反复地检验他的结

论。他发现自杀率在新教国家高于天主教国家；在天主教国家的新教地区高于新教国家的天主教地区，等等。对研究结果的反复验证增加了支持他的结论的证据。

（3）准确性问题

有时质量手段并不能完全正确地反映所研究的变量和概念。因此逻辑推理和重复验证这两个科学原则对保证现存统计资料分析的准确性是极其重要的。流泪并不一定是同情心的正确量度手段，因此假如女人比男人更爱流泪并不能证明她们就一定更富于同情心；把受伤的小鸟放回鸟窝也并不一定是同情心的正确量度；向慈善团体捐款也可能出于同情心之外的其他动机。在这里，没有任何一点可以单独证明女人比男人更富于同情心；但是如果女人在所有这些事情上全都比男人富于同情心，则可以造成支持前述结论的有力证据。

（4）可靠性问题

现存统计资料的可靠性在很大程度上取决于统计资料本身的质量：它们是否精确地报告了事实本身？如果统计资料本身不可靠就会造成严重问题。防止这类问题的第一步就是要看到这个问题存在的

可能性。对数据和报表的性质进行调查，可以使你对数据失真的程度作出估计从而加以校正。如果同时运用逻辑推理和重复验证的方法就能更好地对付这类问题。

（5）现存统计资料的来源

毫无疑问，最有价值的资料汇集是每年由美国商业部出版的《美国统计摘要》。它是美国质量最佳的资料汇集，其中包括各州和大中城市的基本资料。此外商业出版商葛劳塞特（Grosset）和邓洛普（Dunlap）出版的《美国年鉴》在对社会科学研究的可靠和有用方面也是出类拔萃的。联邦机构——劳动部、农业部、交通部等——也不断公布大量的数据资料。

联合国占有大量世界统计资料。它的《人口年鉴》每年提供各国重大的统计资料（出生率、死亡率及其他关于人口的资料）。联合国的其他出版物提供各种分类的数据。

由非政府机构提供的数据数量同样惊人。例如，乔治·盖洛普从1935年起一直不断公布盖洛普民意测验的结果。

在政府各部和大学图书馆，你将发现无数数据在等待你去分析。缺少研究经费并不能阻止出色而

有效的社会研究的进行。

7 历史-比较分析

（1）历史-比较分析举例

社会学一词的创造者奥格斯特·孔德（Auguste Comte）视这一新学科为人类思想发展史的最后阶段。他描绘了一幅人类发展的画卷——从神学阶段到玄学阶段最终发展到科学阶段。他还描绘了科学从生物学及其他自然科学到心理学，最后到科学的社会学的发展过程。

许多晚近的社会科学家也非常注意广泛的历史过程。他们中间有些人认为，社会形式的历史发展过程是从简单到复杂、从以农村和农业为主发展到以城市和工业为主。例如，美国人类学家莱维斯·摩尔根认为，历史上存在着一个从“原始时代”到“野蛮时代”再到“文明时代”的进步过程。现代人类学家罗伯特·瑞德菲尔德（Robert Redfield）描述了从“民俗社会”到“都市社会”的转变。迪尔凯姆认为，社会演变是劳动分工不断发展的过程。卡尔·马克思认为，经济体系有一个从原始形式到封建形式到资本主义形式的历史性进步。在这

个思想指导下，他提出，全部历史是一部阶级斗争的历史——有产者拼命维护他们的既得利益而无产者则为生活得更好而斗争。马克思还预言了社会主义的发展和共产主义的到来。

然而并非所有的历史研究都持这种进化论的观点，有些人主张循环论而不是直线发展论，其代表人是皮蒂瑞姆·索尔金（Pitirim A.Sorokin）。他是俄国革命的参加者，曾任克伦斯基的秘书，失势后开始了他的第二事业：社会学。

卡尔·马克思在对经济史的分析中表达了一种经济决定论的观点。他认为经济因素决定了其他一切社会现象的性质。例如，他认为，欧洲教会的作用是维护资本主义的现状，因此宗教是有权者维护他们对无权者的统治的工具。马克思有一句名言：宗教是人民的鸦片烟。

德国社会学家马克斯·韦伯（Max Weber）不同意这一观点。他并不否认经济因素可以影响其他社会因素，但认为经济并不能解释一切。他认为，经济形式可以从非经济的意识获得。在关于宗教社会学的研究中，韦伯检验了宗教体系作为社会行为原动力而不仅是作为经济条件的反映的程度。他关于这一问题的论述集中在一本题为《新教伦理与资本主义精神》的著作中。

法国神学家约翰·卡尔文(John Calvin)是新教改革的重要人物，他提出，上帝的意愿决定每个人的成功或失败。具有讽刺意味的是，这种观点使他的追随者勤奋节俭、争取成功（特别是经济上的成功）以求证明自己得到上帝的垂青。韦伯认为，卡尔文主义为资本主义的发展提供了一个重要的刺激。卡尔文主义者把钱用于扩大再生产而不是"浪费"在享受上，于是为资本主义的发展提供了必要的资本。

在其他三项研究中，韦伯对犹太教、中国宗教和印度宗教进行了详细的历史分析。他想了解资本主义没有在中国、印度这些古老国度中发展起来的原因。他发现这三种宗教的教义都不能用来促进资本积累和扩大再生产。这些发现从反面证实了他关于新教对资本主义发展所起作用的结论。

（2）历史-比较研究资料的来源

象现存统计资料一样，供历史研究的资料也是取之不尽用之不竭的。首先，关于你要研究的一切，历史学者已有描述。这种"第二手资料"可以为进一步研究打下基础。

人们毕竟更想利用"第一手资料"。当汤玛斯（Thomas）和兹南尼克（Znaniecki）研究本世纪

初叶移居美国的波兰农民适应本地生活的过程时，他们用的是移民给波兰亲友写的家信（他们是从报刊广告栏上找到这些信件的）。有些研究者用旧时的日记作分析。关于家庭生活史的研究，艾伦·罗斯曼（Ellen Rothman）指出下列资料来源：除个人资料外还有许多关于家庭历史的公开记录：报纸上常常登载与家庭的教育、法律和娱乐有关的大量资料；杂志的评论往往能反映家庭生活的一般模式——从如何教育子女到怎样贴糊墙纸；小说故事也可以反映家庭生活的共同经历和观念。

各种组织一般都发行正式文件：宪章、政策声明、领导演说等等。有些组织不但出版周报周刊还编印领导人物讲演集。

政府的官方文件也可为研究提供资料。为研究美国种族关系史，列奥·海金伯萨姆（Leon Higginbocham）查阅了200年来的法律及与种族问题有关的法庭审判记录。

象对现存统计资料应取的态度一样，你不能完全相信这些历史记录，不论是正式的还是非正式的，第一手的还是第二手的。防止出错的办法是重复验证。如果几种不同的资料全都证实了某项事实，其可信性就增强了。如果关于某政治运动的资料全部是从运动内部得来的，你就很难从中得到一个

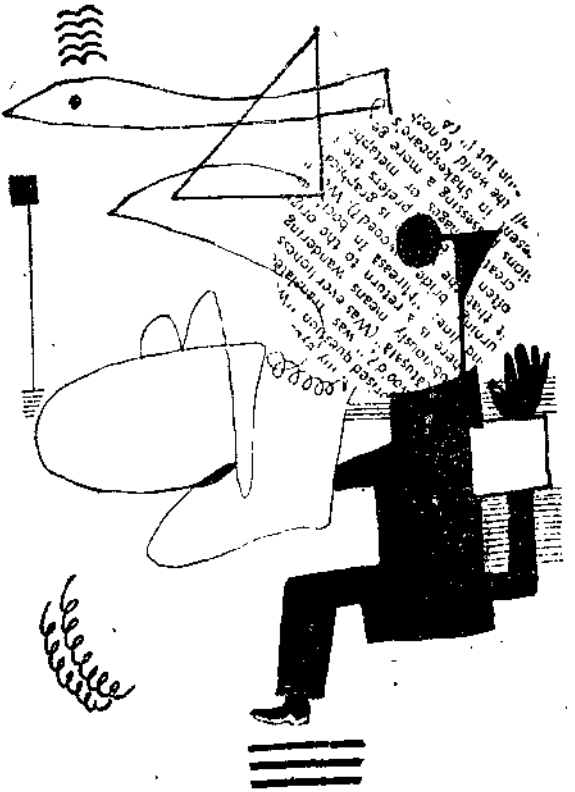
全面的看法。中世纪体面绅士的日记也往往并不能准确反映那个时代的生活。因此在可能的情况下，要尽量从不同的来源收集反映不同观点的资料。

（3）分析技术

由于历史-比较研究属于定性的研究方法，很难将数据分析步骤一一列出。马克斯·韦伯提出"理解"这一方法，意为研究者必须去理解研究对象的环境、观点和感觉，以便正确地解释他们的行为。

历史-比较研究者必须从大量细节中发现模式，如韦伯所谓"理想形式"，即由社会现象的基本特性所构成的概念模式。韦伯在关于科层制的研究中不仅列举了科层制的共同特征，而且对维持科层制正常运转的各种基本因素作了充分的理解，从而创造出科层制的理论模型（理想形式）。

综上所述，社会科学家可以通过多种途径去研究社会生活而并不影响研究对象。



第十二章
评　价　研　究

评价研究（又称方案评价）与其被视为一种研究方法不如说是一种研究目的，其特殊目的在于对社会干预——如一种新教学方法、假释方法或其他方案——的效果作出评价。调查、实验等方法都可以用于评价研究。

评价研究与社会科学一样古老。当人们实行社会改革时，往往需要了解其实际效果，虽然有时这样做是无意识的。近年来，评价研究日益普遍活跃，在社会研究中占据了重要地位。

1 适于采用评价研究方法的研究课题

评价研究方法适用于对计划或实施中的社会干预方案的研究。所谓社会干预是指在特定社会背景

下为制造某种预期效果而采取的行动。在最简单的意义上，评价研究是确定某方案是否达到预期效果的研究。

例如，我们认为允许配偶探监（并允许他们有性生活）是提高囚犯情绪的好方法，并制定了这样的方案。这时，了解囚犯的情绪是否真能因此而有所提高是十分必要的，或许这一作法不但不能提高情绪反而使事情变得更糟。在这种情况下就可以用评价研究来检验这个方案的效果，其中应当包括设计度量情绪高低的尺度，从囚犯中选择控制组和实验组，让实验组的人在一段时间内接受配偶探视，然后比较两个组的情绪高低。假如实验组的人情绪果然提高了，则可能说明改革方案是有效的。

适于采用评价研究方法的研究课题多具有重大的现实意义，不论是对干预方案的研究还是对信仰及价值观的研究都是如此。

2 评价研究的设计

（1）度量尺度

前述度量方法在评价研究中都同样适用，本节将着重论述运用这些度量手段去度量什么的问题。

〔度量效果〕显然，在评价研究中要加以度量的主要变量是效果。假如某项社会方案旨在达到某种效果，你首先必须能够度量这种效果。例如，要降低种族偏见的程度，就应当能度量种族偏见程度；要提高婚姻和谐程度也应当能够对它进行度量。

试举一例，梅耶尔（meyer）和伯迦特（Borgatta）（1959）对精神病某治疗方案效果进行评价研究时曾设计了一套度量“病人痊愈”的指标，其中包括：

①没有再进精神病院

②不再需要恢复治疗

③重建社会关系

④经济独立

⑤能面对现实

⑥身心健全

仅用上述指标度量仍不够具体。健全到什么程度才算身心健全？在现实生活中到底什么叫身心健全？怎样才能知道病人是否已经身心健全了？

〔度量实验的背景〕对与实验方案直接有关的因变量的度量仅仅是开始，对实验背景的质量也很重要。这些变量是实验的外因，但对其影响不可忽略。例如要评价一个训练失业人员使之就业的计划。预期结果的基本度量指标应当是训练结束之后

的就业人数。但是在度量受训人员就业率的同时还应度量整个社会在那个时期的就业率。如果经济不景气，受训人员就业率乍看上去很低，其实结果不能算坏；而有时受训人员全体就业了却不是因为训练班的作用，而是就业机会增多了。由此可见度量实验背景的重要性。

〔度量干预本身〕除了评价实验的结果，还需要对方案中的干预作法本身做出度量，即把人们分为实验组和控制组。如前例，受训人员为实验组，未受训练者为控制组。如果训练班是成功的，你应当发现受训人员比未受训人员的就业率要高。

〔度量其他变量〕对研究对象所属总体进行度量往往是很有必要的，特别是要对方案适用范围作出定义。度量过程包括两个方面：第一，为研究对象的总体下定义，然后在其中随机抽取控制组和实验组。例如在治疗精神病的实验中，应保证控制组和实验组人员的病情基本一致，不要有系统偏差。第二，在度量这些不同变量时，可以用别人已用过的度量尺度，也可以创造新的度量尺度。自己设计度量尺度虽然耗费较多时间金钱，但更准确些。

（2）实验设计

第八章已介绍过多种实验设计方法，其中许多

也适用于评价研究。

例如，我们要对一种治疗阳萎的新方法作出评价。首先应当确定适于参加治疗的病人总体。假定某医院已有一百个阳萎病人，我们可以从现存诊断开始，并保留现存的病情评价。但为了进行评价研究，还需要进一步设计度量阳萎的尺度。如近期内病人是否有过性行为，性交频率，是否达到快感等等。

最简单的设计是把100个病人随机分为实验组和控制组。对实验组病人施行新治疗法，对控制组成员仍使用传统治疗法。实验时间定为两个月。

假定新治疗法包括给病人放映色情影片。对这一实验刺激还要作更具体的规定：多长时间看一次电影？每次看多长时间？单个看还是集体看？放映时医生是否到场？要注意观察病人的表现——是否有些病人看得很热切，有些却心不在焉？应当把全部观察记录下来，积累大量数据供分析用。实验结束后，我们就可以确定是否新治疗法达到了预期的——甚至意料之外的——效果。

这个例子表明在评价研究中应当如何运用第八章论述过的标准研究设计方法。

（3）准实验设计

准实验与“真”实验的基本区别在于它的实验

组和控制组不是随机组成的。在评价研究中，常常不可能得到随机组成的实验组和控制组。在这种情况下，可以运用下列方法做研究设计：

〔时间系列设计〕时间系列设计是关于过程的研究。假定教师创造了一种使学生积极参与课堂讨论的办法：在星期一，班上只有四个学生发言，星期三，老师拿出一部分上课时间让学生讨论一个有争论的问题，结果星期五再上课时有八个学生发言。于是我们可以说，星期三的讨论使学生参与课堂讨论的积极性提高了一倍。（如图12—1）

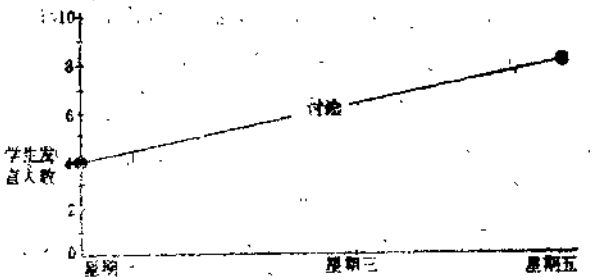


图12—1 对课堂讨论的两个观察：讨论前和讨论后

仅仅星期一和星期五两次观察的比较不足以说明问题。如果有两个班级，一个作为实验组另一个作为控制组，事情要好很多。但在只有一个班的情况下应怎么办呢？可以对学生参加课堂讨论的情况连续作一个学期的观察记录，据此进行时间系列评价。

图12—2是三种可能出现的课堂讨论模式。从哪个模式可以得出星期三的讨论对学生发言积极性有影响的结论呢？显然是第三个模式。

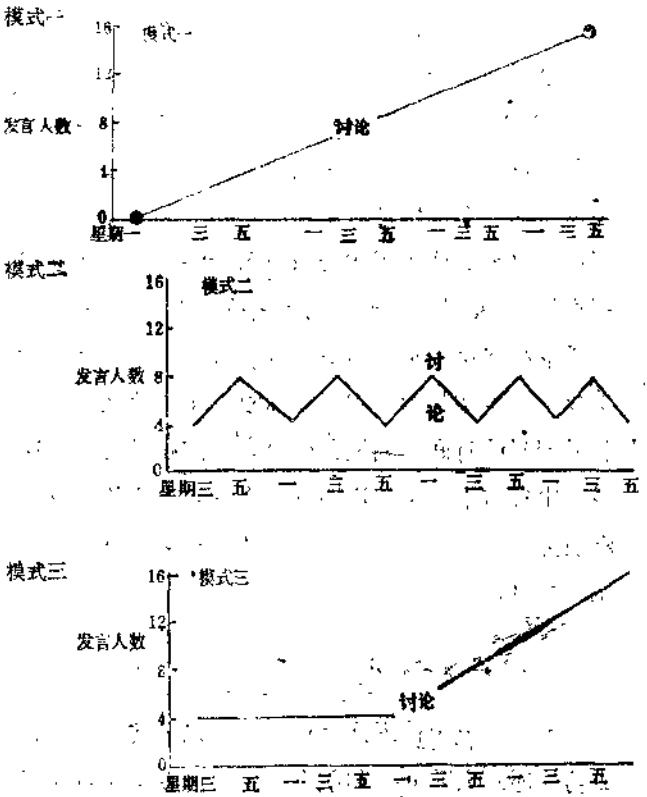


图12—2 从较长时间看课堂讨论的三个模式

〔非对应控制组〕在难以随机抽取实验组和控

制组的情况下还可以利用现成的“控制组”。如在一个班进行外语教学法实验，可以选另一个在年级、性别、民族、智商等方面相去不远的班级作比较。在学期结束时，对使用新教学法和传统教学法的两个班级进行一次相同的外语测验，比较其优劣。

〔多重时间系列设计〕多重时间系列设计是对时间系列设计与非对应控制组设计的进一步发展。此类设计的一个范例是对康涅狄格州限制车速政策的评价。研究者搜集了实行速限前后一段时期的交通死亡报告，发现实行速限后事故确实有所下降。由于交通事故在这段时间内摇摆幅度较大，还不能确定事故下降确实是由于实行了速限。于是他们比较了邻近四州的时间系列统计资料。这四州没有实行速限。比较结果表明，这四州并未发生类似的事故下降趋势。这一比较使实行速限能防止交通事故的结论更加可信。

（4）成败之标准

评价研究中最艰巨的任务在于确定某个方案是成功的还是失败的。如前例，新外语教学法可能提高了学生成绩，但提高到什么程度才算“成功”？配偶探监方案可能提高了囚犯的情绪，但提高到什么程度才能认为整个方案是成功的？

显而易见，这里不可能有绝对的答案。这个难题就导致了代价-利益分析方法的出现。为某一方案付出的代价与它带来的利益是否相当？如利益高于代价则应继续实行，反之则应放弃那个方案。如果你用20元代价制造的产品只能卖出18元钱，为什么还要去造它？遗憾的是在评价研究中很少会遇到如此简单的情况。新外语教学法的实行使学校当局在每个学生身上多花了100元钱，学生平均分数提高了15分。由于考试分数不能换算成钱数，还是难以确定代价与利益之比究竟如何。

在现实生活中，成功与失败的尺度可以从几种方案的比较中获得。假如另一种外语教学法只需要在每个学生身上花 50 元钱却能将平均分数提高 20 分，它显然比前述教学法更成功些。

综上所述，成败标准常常是一种约定。负责实行某方案的人可以事先确定一个预期的结果作为成功的标志。

3 某些有关的社会因素

（1）“后勤”问题

作为一个军事用语，后勤是指食物枪枝等等物

资的供给。在评价研究中使用这个词则是指如何获得研究所需的一切必要条件。这个问题往往比人们预料的要严重得多。

以前面提到的配偶探监方案为例。假定在实行方案的第四天，一个男囚犯打倒他的妻子，换上她的衣服逃出监狱，这下你的研究就会大大地复杂化。也许典狱长决定停止实验；或者他可能会检查由你根据随机原则抽出的实验组成员，指出让某人留在实验组是"冒险"；或者他决定对实验组的囚犯加强警戒，面这个措施可能大大破坏接受探视的囚犯的情绪。这时，实验刺激的内容实际上改变了。实验结束后，有些数据是改变前的最初刺激的反映，其他则是改变后的刺激的反映。虽然你可以把它们分开，但你精心设计的研究被搞得乱七八糟。

（2）伦理问题

评价往往同伦理问题纠缠在一起，有时是社会干预的作法本身引起伦理争论。例如一个关于对小学生进行性教育方案的评价研究不可能不涉及对性教育这件事本身的激烈争论。评价研究设计要求某些小学生接受性教育，当这些孩子的家长因此怒不可遏时，研究人员就要负直接责任。

再看"光明面"：也许实验方案对参加者大为

有利，如一个可以减少工伤事故的安全生产方案。那些分在控制组因而被剥夺了这种利益的人该怎么想？评价研究者的作法将被视为控制组成员发生工伤事故的部分原因。

讨论这些问题并不是为了给评价研究遮上一层阴影，而是为了使评价研究者看到他们的行为在现实生活中可能发生的效果。第十九章对这一问题还要详加论述。

（3）研究结果的应用

评价研究在现实生活中遇到的另一个问题是：认定某种作法的成败优劣并不等于能决定其命运。例如，尼克松总统曾设立一个特别全国委员会专门研究淫秽物品造成的社会后果。经过认真的、多方面的评价研究，这个委员会得出下列结论：淫秽物品并没有造成那些常常归咎于它的反面社会后果。如，淫秽物品并未增加性犯罪的可能性等等。看到这种研究结果，你也许期望立法会因此更自由化一些，但实际情况却是总统说委员会搞错了。这是一个很有戏剧性的例子。类似的事情还有很多。

评价研究结果不是总能被采用的原因有三：第一，有时，评价研究结果没有使用一般人能够理解的方式加以表达。第二，评价研究结果有时与人们深信

不疑的事情相反。如前述关于淫秽物品的例子。如果人人都认为淫秽物品会导致性异常行为，则与之相反的研究结果很难马上被人接受。由于同样原因，当哥白尼说地球绕太阳旋转时人们认为他是发了疯。谁都知道地球是纹丝不动的嘛！应用评价研究结论的第三个障碍是既得利益。假如某人设计了一个改造罪犯的方案并已被公认可以防止犯人重新犯罪，而你的评价研究却发现这个方案根本没有作用。显然，你的研究结论很可能遭到方案设计者的激烈反对并被说得一文不值。

4 社会指标研究

另一个正在疾速发展起来的社会研究领域是社会指标的研究。正如经济学家运用人均国民生产总值这类数字作为国家经济发展的指标那样，社会学家可以用同样方法对社会的各个方面作出检验。

如果我们想比较不同社会为其居民提供的保健条件，则可以将其死亡率（每千人每年死亡人数）或婴儿死亡率（每千个一岁以下儿童死亡数）加以比较。此外，还可以设计多种其他度量尺度：人均拥有医生数；人均拥有病床数；人均住院天数等等。

（1）死刑及其威慑作用

死刑是否对死罪如谋杀有威慑作用？每当某个州考虑取消或恢复死刑时，这个问题总要引起激烈的争论。支持死刑的人争辩说，对死刑的恐惧可以防止潜在的谋杀者动手杀人；反对死刑的人则坚持认为死刑并不存在这种威慑作用。对社会指标的研究可以用来解答这个问题。

假如死刑对谋杀有威慑作用，谋杀率在实行死刑的州就应当低于那些取消了死刑的州。威廉·贝利（William Bailey）的数据（表12—1）与这种推测截然相反：谋杀率在实行死刑的州高于取消死刑的州。

表12—1　　有死刑州与无死刑州谋杀发案率
（1967，1968）（1：100,000人）

	无死刑州		有死刑州	
	1967	1968	1967	1968
一等谋杀案	0.18	0.21	0.47	0.66
二等谋杀案	0.30	0.43	0.92	1.03
谋杀案总计	0.48	0.64	1.38	1.59

对贝利的数据可能做出另一种解释，即选择死刑是果，高谋杀率是因：那些谋杀率较高的州选择了死刑，谋杀率较低的州取消了死刑。即使如此，

实行死刑应当使谋杀率降低，取消死刑应当使谋杀率增加。但事实并不是这样。时间系列分析表明，取消死刑并未使谋杀率增加；实行死刑也未使谋杀率降低。

当前，社会指标的研究包括两个方面：一方面，进一步发展和完善那些可以最有效地度量社会生活的指标；另一方面，进一步发现社会生活中各种变量之间的关系。

（2）计算机模拟

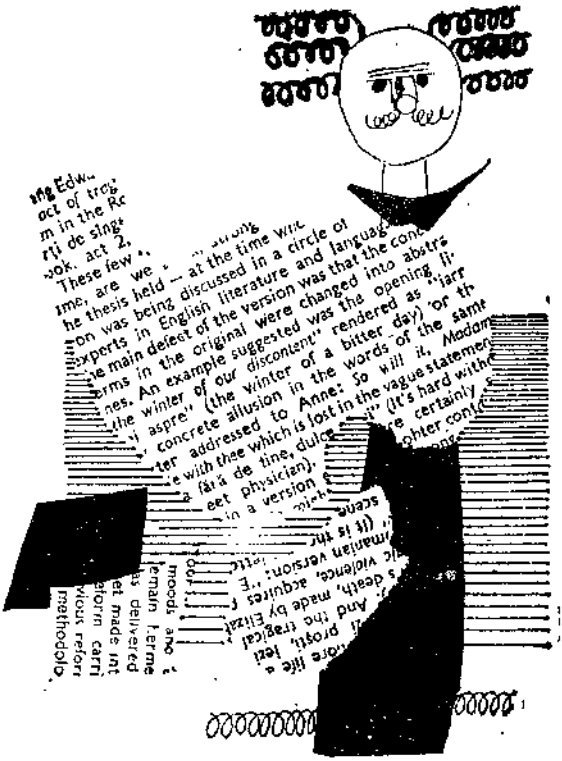
社会指标研究的一个更吸引人的前景是计算机模拟研究的发展。我们把那些描述社会变量间关系（如人口增长与汽车数量增长之间的关系）的数学方程式输入计算机。当存贮了大量这类方程式之后，就有可能通过计算机而不是实际生活来检验某种社会变迁的后果了。

例如，某州打算使旅游业翻一番，可以把它输入计算机模拟模式，并在几分钟内得到对由旅游业增加而导致的直接间接后果的描述，诸如应当建立哪些新的公共设施，哪些机构（如警察、救火队等）应当扩大及扩大幅度，就业人数将增加多少，应当提供哪些相应的职业培训，居民收入和税收将增加多少等等。

计算机模拟的一个范例是多内拉（Donella）和丹尼斯·麦多斯（Dennis Meadows）及其同事对地球上各类自然资源（如石油、煤、铁等）储备量的研究。社会指标使我们能在巨大的规模上作出决策，用计算机模拟来展示某项政策可能的前景，从而避免在实行中造成真正的损失。

第四篇　数据分析

在第四部分的各章中我们将讨论数据分析这一介于观察与研究报告之间的步骤。



第十三章
资料之数量化

本章将讨论如何将社会科学资料转化为适于进行数量分析和可被计算机读入的形式。

1 关于计算机

在开始讨论之前，首先应充分意识到我们目前正处于数据处理革命的过程之中。在社会科学中，打孔资料卡将很快被磁带存贮等技术所淘汰。这里之所以介绍打孔资料卡技术，部分是由于有些人仍在使用它，此外，通过它能够说明数据档案结构的逻辑。

（1）旧式计算机

IBM打孔资料卡共有80列，12行。存贮数据的

方法是在这些行列中打孔。打孔机的用法与打字机相似。数据必须转化为可被计算机读入的形式。例如，实验对象的性别可以被记录在卡片的第5列。如果某研究对象是男性，他的卡片第5列就打1；女性则打2。研究对象的年龄可以记录在第6、7两列，如研究对象是35岁，则这两列分别打为3和5。也可以将年龄分类贮存在一列中，如1代表低于20岁，2代表20—29岁，等等。

每张卡片代表一个研究对象（一个分析单位）所占有的全部数据。如分析单位是报刊评论，则每张片卡代表一篇评论。例如，其中两列可以用来记录这篇评论发表年份的最后两位数字。

回到图9—7，问题1标明“(6—15)”意为对问题a—j的回答将存贮在打孔资料卡的第6—15列。答案下面的数字（1—5）1代表“强烈赞成”，2代表“赞成”，余类推。

（2）新型计算机

新型计算机可以处理更加复杂的分析程序。首先，它可以超越简单计数和分类的功能，同时运用几种变量和多项统计值去计算和提供更加复杂的结果。其次，使用存贮在磁带上的数据远比存贮在打孔资料卡上的数据来得方便迅速，计算起来也快得

多。

目前，多数计算机分析是在计算机主机上输入编好的软件（如SPSS）。你可以带着自己的数据（打孔资料卡或磁带）到计算机中心去，或者提取存贮在计算机中心数据库中的某组数据。输入数据和指令即可得到计算结果。

近年来又发展了两项新技术。一是在计算机中心外设立远距离工作输入系统。研究者可以不去计算机中心而在各自的工作地点通过计算机终端把编好的程序输入中心。另一项技术更加方便，可以通过计算机终端随编程序随输入，随时得到结果，而不用事先编好全部程序输入。手提式的终端可以带回家或携带旅行，使用时只需用一般电话把它与计算机中心联系起来即可。由此可以建立全国的计算机网，使研究者可以坐在加利福利亚家中，得到马萨诸塞州某计算机存贮的一组数据，用得克萨斯某计算机存贮的程序分析这些数据，最后分析结果回到加利福利亚的使用者手中同时存入他在密执安某同事的计算机以备进一步检定。

最新的发展是微型计算机的应用。它体积很小，同打字机相似，运算过程可由一般电视机显示，用直径五英吋的圆片型磁带存贮数据，附带一小型打印机。由于微机的便宜、可靠、轻便和容易

使用，它将会大大发展起来。

2 数据处理术语

〔个案〕个案是指研究中的一个具体研究对象。例如，对于调查来说，一个访问对象就是一个个案；对于流行歌曲内容分析来说，一首歌曲就是一个个案；对于实验来说，一个实验对象也就是一个个案。个案的数量（通常用字母N表示）就是调查对象的数目、实验对象的数目等等。

〔档案〕档案是属于某一特定个案的全部数据的集合。因此，关于某一调查对象的全部信息就组成这个人的档案；关于某个实验对象在实验过程中的全部经历和行为就组成了这个实验对象的档案。

3 编　录

为了进行定量的分析必须把资料转化为数字形式并将其存入数据档案中一定的位置。把资料转化为数字形式的过程称为编录。第十一章内容分析方法关于编录过程的讨论与本节讨论的方法很近

似。与内容分析的编录过程不同的是，运用其他研究方法时常常需要在资料收集上来之后再开始编录。例如，开放型问题的答案是非数字形式的，因此在分析之前必须对其进行编录。实地研究者对定性形式的笔记进行定量分析也需要对其进行编录。

编录的目的是把杂乱的信息归纳为一组由有限的属性组成的变量。例如，研究者向调查对象询问其职业，答案的种类会有很多。虽然可以为每个职业标一个不同的数字，但这种作法不会对分析有多少助益，因为分析需要建立在某些对象所共有的性质之上。

职业这一变量有几种现成的编录方法（虽然都不是很完善）。一种方法是将职业分为专业管理人员、职员、半熟练工等；另一种是按经济中的不同行业区分为产业、卫生、教育、商业等等；还有把这两种方法结合在一起用的。

用哪种方法编录应依研究的理论概念而定。对有些研究来说，把所有的职业编录为白领和蓝领两大类就够了；对另一些研究来说，可以编录为自我雇用和非自我雇用两类。研究和平问题的人只想了解某个职业是否与国防有关。

虽然编录方法应当视分析的特殊要求而定，但还是有一般规律可循的。例如，开始分类细一些，

当分析不要求过细分类时可以合并；但如果一开始分类很粗，分析时有需要却不能再往细里分了。因此，分类应遵循宁细勿粗的原则。

（1）编录类别的设计

编录的基本方法有两种。第一种是在研究之初就设计出相当完备的编码标准，这个标准当然是以研究目的为基础的。如前例，研究和平问题就可以根据是否与国防有关来编录职业。假定你对一个正在形成过程中的新教派作参与研究，想了解新教徒入教的原因。研究过程中你得到一个印象：新教徒似乎把这个教会视为家庭的代替，于是把编录类别定为对这一宗教有这种看法和没有这种看法的两大类，进一步把研究对象的家庭状况编录为有家庭和没有家庭两类。你发现，没有家庭的人比有家庭的人更倾向于视这个教会为家庭的代替。这里的一个潜在危险是，每当你发现一个研究对象是没有家庭的，你会不知不觉地去寻找他视教会为家庭的证据。因此在可能的条件下，应当请其他人来编录同一内容，看结果是否相同。

第二种方法是先记录各种答案然后分类。如对某个开放型问题的答案可能有50或100种，可以把它们先全部开列下来，然后把答案相近的归为一类。

编录类别同变量的属性或封闭型问题的答案一样，必须是穷尽的和相互不包容的。每个信息必须能够归入一个且只能归入一个类别。问题在于那些适于归入两个类别或哪个类别也不适于归入的答案。

（2）几点建议

在研究对象做了一个以上的选择，如他既选择“强烈赞成”又选择“赞成”时，不应当同时记录两项，而可以按照一种统一的逻辑优先原则，如凡是这类情况都记为“强烈赞成。”

有一种情况需要把某一对对象归入一个以上的类别。如在一个实验中，你想分析研究对象对困难的反应愤怒、忧愁、幽默、暴躁、放弃努力、勤奋等等。你发现多数研究对象表示出不只一种情绪。这种情况下有两种处理方法。

第一，可以为每种反应单独设立一个变量，记录这种反应是否出现。如设第28列为愤怒，用编录做字1表示研究对象表现了愤怒，2表示没有愤怒；设第29列为忧愁，依此类推。

第二种方法是对某一问题只设对象可能给出的答案种类的列数。假如每个对象录多只做出了三种反应就设三列。数字1代表愤怒，2为忧愁，3为幽

默，4为暴躁等等。如果某个研究对象表示出愤怒和忧愁，则在第一列输入1，第二列输入2，第三列输入0表示无；另一对象上表现出暴躁，则在第一列输入4，其他两列为0。为什么不可以在0的位置留下空白呢？因为第一，事后你可能会分不清它是0还是输入遗漏；第二，有些计算机的空白并不等于0而另有意义，留空白会造成混乱，甚至使整个计算过程夭折。

4 编录表的设计

编录过程的最终结果是将资料转化为数字。这些数字代表了变量的不同属性，这些变量又在数据档案中占有一定的位置。编录表就是对变量位置及数字意义的描述。编录表有两项基本功能：第一，作为编录过程的指南；第二，使研究者在分析过程中便于查找变量及各个数字所代表的意义。图13—1是第九章提到的问卷所使用的编录表。

卡片1 编录表

列	内容
1 4	研究对象的号码
5	卡片号码：1
6	A-1a “联合国应当成为世界政府。”

1、强烈赞成
2、赞成
3、反对
4、强烈反对
5、未详
0、无回答

7 A-1b "侮辱国旗的人应当被送送监狱。"
1、强烈赞成
2、赞成
3、反对
4、强烈反对
5、未详
0、无回答

卡片3 编录表

列	内容
1-4	研究对象的号码
5	卡片号码：3
6	B-9 "你认为老师对你课堂发言的评价重要吗？"
	1、非常重要
	2、重要
	3、不太重要
	4、根本不重要
	0、无回答
7	B-10 "你认为得好分数重要吗？"
	1、非常重要
	2、重要
	3、不太重要
	4、根本不重要
	0 无回答

卡片5　编录表

列	内容
1-4	研究对象的号码
5	卡片号码：5
	E-1 “大学期间，你参加下列活动吗？”
6	E-1a “校队体育运动”
	1、很多
	2、一些
	3、很少
	4、没有
	5、无回答
7	E-1b “校内一般体育运动”
	1、很多
	2、一些
	3、很少
	4、没有
	5、无回答
8	E-1c “社会活动或政治团体”
	1、很多
	2、一些
	3、很少
	4、没有
	5、无回答

图13—1　编录表举例（片断）

每张卡片的第1—4列都是研究对象的号码，第5列是卡片号码。如果我们想了解学生参加社会活或政治团体的情况，图13—1的编录表表明这一信息

记录在第5张卡片的第8列上。所有在这一列标明为1的研究对象是报告参加活动“很多”的人，数字为2的是参加了“一些”活动的人，余类推。

5 编录与数据输入的几种选择

有几种方法能够使编录与数据输入过程结合起来。这里介绍其中四种不同的选择。

（1）转换表

数据处理的一个传统方法是把数据誊写在转换表上，这种表格有80列，与打孔资料卡的列数相同。编录人把应当输入的数字写在适当的列上，打孔员就可以按照转换表输入数据了。

（2）页边编码

这种方法不必使用转换表，而是在问卷的页边上直接编码，使数据与打孔资料卡上的列数一一对应。

（3）编辑输入

这种方法适用于篇幅过长的问卷。为了直接从

问卷上输入数据，应当在设计问卷时就标明列数和每种答案的数字。但使用这种方法之前一定要对填好的问卷进行编辑。要检查是否每个问题都回答了（在未回答的问题后注上0或其他标准数字）；是否有的问题出现一个以上的答案（根据统一的逻辑优先原则把它们改为一个答案）；及其他不够清楚的地方。封闭型问题的答案可以直接进行数据输入；开放型问题的答案则需要重新编录然后输入。

（4）直接输入

一种更先进的数据处理系统，即直接输入的方法正在普及。编录人面对显示编录表的计算机终端屏幕，检验全部问卷，确定适当的编录类别，然后直接将数字输入计算机。这种方法省去了使用打孔资料卡、转换表以及从编录员到打孔员费两道手续的麻烦。

6 数据之清理

不管你多么认真，总是难免出错。根据不同的数据处理方法，错误有可能来自编录误差或输入误差。数据的清理主要有两种方法，一种是可能性清

理法，另一种是逻辑性清理法。

（1）可能性清理法

任何变量只有一组有限的属性。如对性别这一变量的填写只有三种可能性：数字1为男，2为女，3为无答案。如果在清理数据过程中你发现某人性别数字为7，那显然是输入错误。

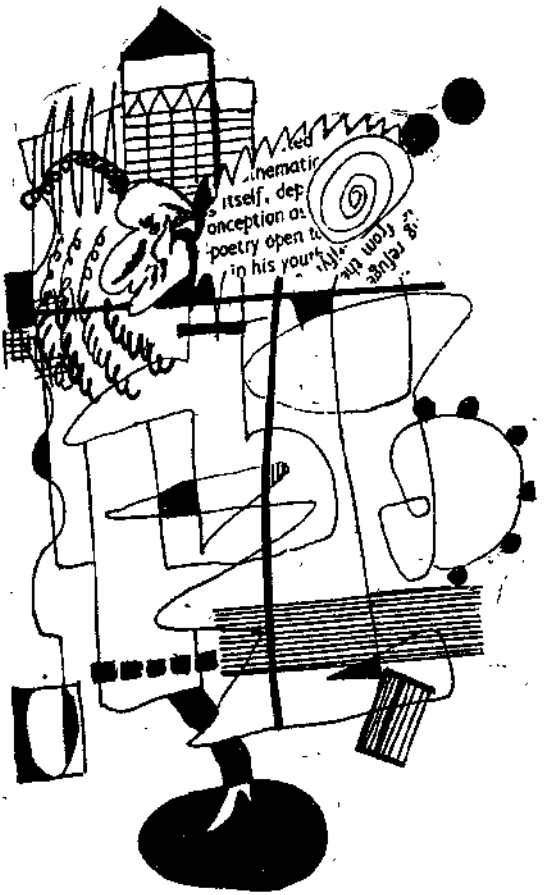
可能性清理可以利用计算机来进行。把全部可能出现的数字输入计算机，它就会查出不应当出现的数字。

（2）逻辑性清理法

逻辑性清理法较为复杂。数据的逻辑结构可能对某些研究对象作出限制。如某问卷问到生育数字。所有的女性研究对象都应当作出回答（可设计一个特殊的数字代表无答案者）；而所有的男性都不应当回答。如果某男性研究对象的数据表明他生过三个孩子，显然是输入错误。

逻辑性清理法也可以利用计算机进行，但比可能性清理法复杂。在计算机上要使用“如果怎样则怎样”的程序语句。但有时也有捷径，如有些男性被错误地赋予了母亲身份，在分析时只要只用女性数据就可以避免这种误差。当然，这个建议并不是为

草率行事提供借口，“不干净”的数据总是会把研究引入歧途的。



第十四章
基本分析方法

1 单变量分析

顾名思义，单变量分析是在一个时间点上对某一个变量的分析。下面将讨论单变量数据分析的逻辑和形式。

（1）分 布

描述单变量数据的最基本形式是描述全部单个个案的情况，即开列每个研究对象的变量取值。假定你想了解公司经理的年龄，最直接的方式是开列这些年龄：63岁，57岁，49岁，62岁，80岁，75岁，55岁等等。这种作法可以为读者提供数据的全部细节，但缺点是不便于做任何进一步的数据处理。

一个可以把数据安排成便于处理的形式同时不致于失去任何细节的方法是描述全部数据中有5位

经理是38岁，7位39岁，18位40岁等等。这种方法可以避免重复地写出每一数据。

一种更易进行处理但失去了部分细节的方法是描述分组数据的频数分布，如：246位经理低于45岁，517位在45到50岁之间，等等。

此外还可以用比例数来描述数据的分布。例如可以说，百分之几的经理年龄在45岁以下，百分之几在45到50岁之间，等等。（见表14—1）计算比

表14—1　　单变量分析一例

经理年龄（假设）	
35岁以下	9%
36—45岁	21%
46—55岁	45%
56—65岁	19%
66岁以上	6%
	100%＝（483）
	（40）

例需要确定基数，代表百分之百的数字。问题在于如何处理未详数据。例如在某研究中，有若干人没填年龄，对此有两种处理方法：第一，可以仍以全部研究对象的总数为基数计算比例，同时写明未填年龄的人所占比例。第二，以对问卷作了回答的人为基数计算比例，同时标明未填年龄的人的数目。

但不用计算其比例。

如何选择应依研究目的而定。如果你想比较样本年龄分布和总体年龄分布，最好不把“未详”数据算进基数，因为“未详”不是一个有意义的年龄组，把它计入基数会搞乱样本总体的比较。

（2）集中趋势

在说明频数、比例之外还可用平均数或集中趋势来描述数据，即众数（出现频率最高的取值），算术平均数和中位数（位于按顺序排列的全部取值中心的取值）。

假定我们做一实验，要求研究对象在一定时间内数成某事。有些人第一次尝试就做成了，有些人第二次才做成，还有些人需要更多的时间。下列实验数据表明，7个研究对象第一次就数成了，12个人需要做第二次，余类推。

尝试次数	成功人数
1	7
2	12
3	22
4	18
5	6
6	2

众数最容易发现。由于三次做成的人数最多，众数就是3。但众数用处不大，在研究报告中提到众数的极为少见。

平均数是先求出全体研究对象尝试次数的总和：（7×1）+（12×2）+（22×3）+（18×4）+(6×5)+(2×6)=211。然后用实验人数（67）去除：211/67=3.15。

中位数就是“中间”值。将67个研究对象按尝试次数多少排列，第34个人是位居中间的人，他是三次数成的，所以中位数是3。

研究报告往往要计算和说明平均数和中位数。使用平均数时要注意，它对极端值非常敏感。一组数据中极大或极小的数对平均数影响比对中位数要大。

（3）离散趋势

仅有平均数不足以使未见到原始数据的人了解全部数据的情况，还必须描述离散趋势。度量离散趋势的最简单尺度是全距，即从最高值到最低值的距离。因此，当你报告一组研究对象的平均年龄为35岁时，还应报告他们的年龄幅度是从18岁到69岁。标准差也是度量离散趋势的方法，但较为复杂。如第五章关于抽样分布标准误差的例子，研究报告

说研究对象年龄的标准差为10.5岁，意为68%的研究对象的年龄不超过平均岁数正负10.5岁的范围，即24.5岁至45.5岁之间。

还有其他度量离散趋势的办法。如报告智商数时常用四分位间距，分数最高的四分之一人得分间距，分数次高的四分之一人得分间距等等。如果分数最高的四分之一人得分间距是从120到150；分数最低的四分之一人得分间距是从60到90，则得到平均数（假定是102）的四分位间距是从120到90，即30。

（4）连续变量与离散变量

上述方法并非对所有的变量都适用。一般来说，变量有两种：连续变量和离散变量。年龄是连续变量，而不象性别或军阶一类离散变量那样从一个类别跳到另一个类别。如果分析离散变量（如定名或定序变量），上述的一些方法就不适用了。严格说来，中位数和平均数只适用于定距或定比变量。例如对性别这一变量只能用众数来描述，报告性别的平均数、中位数或离散趋势就不正确了。

2 分组比较

我们还可以对研究对象进行分组描述，如表14—2描述了男工和女工的收入状况，同时提供了女工收入中位数对男工收入中位数的比率，表明女工的收入是男工的一半多一点。

研究者做分组描述有时纯为描述，但更经常的是为了比较：女工收入比男工少。如果将白人与黑人的收入作比较，也会发现类似的情况。这种分析建立在因果关系之上，自变量导致因变量，性别影

表14—2　　男女工人收入中位数：1967—1977

年代	女工	男工	女/男收入比率
1977	$8,618	$14,626	0.59
1976	8,622	14,323	0.60
1975	8,449	14,175	0.60
1974	8,565	14,578	0.59
1973	8,639	15,264	0.57
1972	8,551	14,778	0.58
1971	8,369	14,064	0.61
1970	8,307	13,993	0.59
1969	8,227	13,970	0.59
1968	7,763	13,349	0.58
1967	7,503	13,021	0.58

响收入。

3 双变量分析

与单变量分析不同，分组比较提供了包括两个变量的分析，又称双变量分析。如前所述，单变量分析的目的纯粹为描述。分组比较的主要目的仍在于描述，但增加了比较的因素。在社会研究中，大多数双变量分析又增加了另一种因素：变量之间的关系。

从某种角度看，表14—3可以是分组比较，但从另一种角度看，它是解释性双变量分析。它表明变量性别影响男女平等观点这一变量。观点是因变量，它部分地决定于性别这一自变量。我们不再以男性和女性为不同的分组，而视性别为对其他变量

表14—3　　对“你赞成还是反对男女平等”这一问题的回答

	男	女
赞成	83%	75%
反对	17%	25%
	100%	100%
	(400)	(400)
无答案	(12)	(5)

有影响的变量。

表14-3表明，女性比男性更赞成男女平等（女性的75%；男性的63%）。比例在表上是横向计算还是纵向计算是人为规定的。纵向计算时，列的合计为百分之百；横向计算时，行的合计为百分之百。

制作解释性双变量表的步骤如下：

第一，将研究对象按自变量的不同属性分组。

第二，用因变量的不同属性对这些组作出描述。

第三，比较自变量各组在因变量属性上的差异。

以分析性别与男女平等观点的关系为例：

第一，将研究对象分为男性组和女性组。

第二，用赞成还是反对男女平等对这两个组作出描述。

第三，比较男性组和女性组赞成男女平等比率的差异。

4 多变量分析简介

本书第十六、十七两章将详细讨论多变量分析。这里只简单介绍一下多变量表：由两个以上变量构成的表。

制作多变量表的步骤与制造双变量表相似。但

双变量表只有一个自变量和一个因变量，多变量表有一个以上的自变量来解释因变量。

仍用前例，假定我们认为年龄也影响人们在男女平等问题上的观点：年轻人比老年人更赞成男女平等。做表的第一步是把整个样本按两个自变量的全部不同属性分组：年轻男性、老年男性、年轻女性和老年女性。然后用因变量的不同属性来描述这些组，并进行比较（见表14—4）。

表14—4 多变量关系：观点、性别和年龄（假设）"你赞成还是反对男女平等？"

	30岁以下		30岁以上	
	女	男	女	男
赞成	90%	78%	60%	48%
反对	10%	22%	40%	52%
	100%	100%	100%	100%
	(200)	(200)	(200)	(200)
无答案	(2)	(10)	(9)	(2)

从这个表可以得出下列结论：

第一，不论男女，年轻人都比老年人更赞成男女平等。只看女性，赞成男女平等的在30岁以下占90%，30岁以上占60%。

第二，不论年轻年老，女性都比男性更赞成男女平等。只看30岁以下的人，赞成男女平等的在女性中占90%，在男性中只占78%。

第三，年龄对观点的影响比性别的影响要大。年龄的影响造成30%的差异（90%:60%，78%:48%……）；性别只造成12%的差异（90%:78%，60%:48%……）。

第四，年龄和性别对观点的影响是相互独立的。在一个自变量的某一属性之内，另一个自变量的不同属性对观点的影响仍然存在。

第五，年龄和性别对观点的影响又是相互累积的。年轻女性是最赞成男女平等观点的，而老年男性是最反对男女平等观点的。

用同样的数据还可以采用另一种格式作表。赞成的是90%，反对的自然是10%，所以可以略去反对的比例，如表14—5。

表14—5 “你赞成还是反对男女平等？”（表14—4之简化）赞成者所占比例

	女	男
30岁以下	90%	78%
	（200）	（200）
30岁以上	60%	48%
	（200）	（200）

表14—5表明两个自变量中不同部分赞成男女平等的比例，括号里的数字是研究对象人数。如，样本中30岁以下的女性有200人，其中90%赞成男女

平等，即180人赞成，其他20人（10%）反对。

最后介绍一个更加复杂的多变量表作为本章的结束。这个例子是社会学家莫瑞斯·罗森伯格（morris Rosenberg）所作的关于男少年自我评价的研究。如表14—6，罗森伯格同时检验了三个变量对自我评价的影响：（1）兄弟姐妹中的性别分布：研究对象家庭中是女多男少，男多女少，还是男女数目相等？（2）排行：研究对象在家里排行第

14—6 男少年自我评价与性别分布、排行及学习成绩的关系

自我评价	研究对象排行在前一半或中间 学习成绩 A—B	C	D—F	研究对象排行在后一半 学习成绩 A—B	C	D—F
			家中女多男少			
高	56%	45%	41%	46%	60%	64%
中	20%	27%	27%	19%	18%	18%
低	24%	28%	32%	35%	22%	18%
总计	100%	100%	100%	100%	100%	100%
人数	(79)	(104)	(41)	(26)	(65)	(22)
			家中男多女少或男女数目相同			
高	51%	46%	29%	42%	44%	30%
中	26%	27%	18%	32%	33%	20%
低	23%	32%	53%	26%	23%	50%
总计	100%	99%	100%	100%	100%	100%
人数	(168)	(240)	(102)	(78)	(96)	(56)

几？（3）学习成绩：研究对象在学校成绩如何？

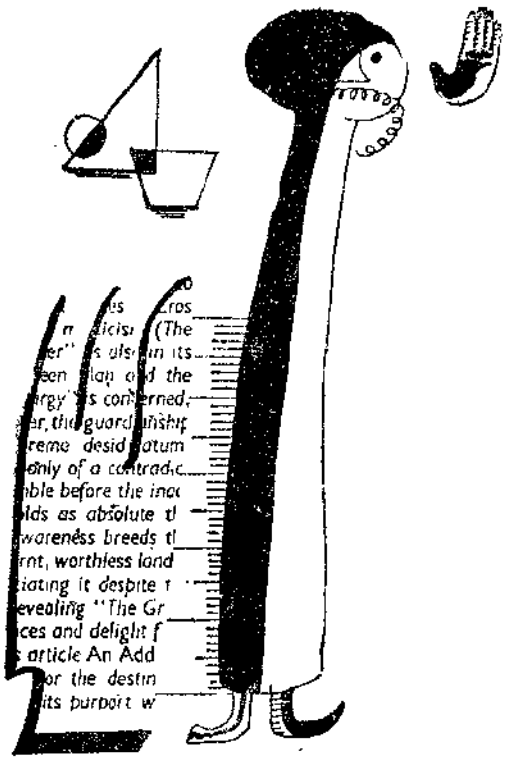
让我们看看表14—6提供了哪些信息。为简便起见，我们只看自我评价高的男少年。要了解兄弟姐妹中性别分布对自我评价的影响，应当比较表的上下两部分的第一行：56%比51%，45%比40%，41%比29%，46%比42%，60%比44%，64%比30%。你发现模式了吗？在其他一切条件相似的情况下，男少年在女多男少的家庭（上半部）比在男多女少的家庭自我评价更高。一个更加细微的差别是，这种情况在成绩差的男少年中更为明显。在成绩好的男少年中这种差别很小。

排行的影响怎样？这次我们比较表的左右两部分的第一行：56%比46%，45%比60%，41%比64%，51%比42%，40%比44%，29%比30%。在其他一切条件相似的情况下，比较排行的影响。这次很难看出什么模式。有时排行大的男少年自我评价高，有时排行小的自我评价高。

最后，学习成绩对自我评价有何影响？我们的想象可能是成绩好的男少年比成绩差的男少年自我评价更高。让我们来看事实。比较全表四个部分的第一行：56%比45%比41%，46%比60%比64%，51%比40%比29%，42%比44%比30%。我们的想象是否与数据相符呢？学习成绩在排行大的男少年

中似乎比在排行小的男少年中对自我评价有更大的影响。对此罗森伯格提出了一种解释：排行小的男少年属于一种独特的自我评价类型，即无条件的自爱。由于父母和哥哥姐姐的宠爱，排行小的男孩的自我评价往往不受学习成绩好坏的影响。

这个表提供了运用比例表做多变量分析的全貌。可以说在一个较浅的层次上，你已经看到了数据分析的全过程。



第十五章
指标、尺度和分类

本章是前面关于概念化（第五章）和实施化（第六章）讨论的延续。前面讨论的是社会研究者为一项研究设计度量尺度的方法；本章将描述怎样将它们运用于数据分析。

在社会科学研究中往往需要设计组合的度量尺度，原因在于：第一，很难用单一的标准来度量复杂的概念，特别是观点和倾向这类复杂的变量。如，很难用研究对象对问卷中某个问题的回答度量其种族偏见程度、或宗教信仰程度、或政治倾向、或异化程度等等。应当设计若干问题，每个问题代表变量的一部分，而没有任何一个问题可以单独描述这一变量。

第二，如果你想运用定序量度处理变量，例如将研究对象按异化程度的高低加以排列，单一的标准很难提供足够的类别，而由几项内容组成的指标

和尺度却可以做到这一点。

此外，单一的标准只能提供某变量的大概情况，几种数据的组合则可以提供更全面更准确的情况。指标和尺度是一种数据简化设计：将几种标准概括为一个数字同时保留对各个标准内容的详细描述。

1 指标与尺度

指标与尺度这两个用语在社会科学研究中常常不是严格区分开的，有时甚至可以混用。在区分它们之前，先来看看它们一致的地方。

指标和尺度一般都是对变量的定序量度，即在某个变量上给人(或其他分析单位)排列顺序，如宗教信仰程度、异化程度、社会经济地位、种族偏见程度及智力商数等等。某人宗教信仰程度的指标分数（或尺度分数）表明其相对于其他人的信仰程度。本书讨论的尺度和指标都是变量的组合量度：包括一个以上标准的量度。如某人的智商分数是根据他对大量问题的回答得出的。

本书将通过评分方式的不同来区别指标和尺度。指标由单个标准分数的简单累计组成，尺度的

组成测可利用标准自身的趋强结构。

试举一例。假定我们想度量人们对共产主义者公民权的观点。在问卷中可以提出这样一个问题：你认为共产主义者是否应当被允许从事下列职业（1）律师，（2）医生，（3）政府部长，（4）工程师，（5）记者。有人认为五种职业都可以，有人认为都不可以，还有人认为其中有些可以有些不可以。这些职业并没有按其重要性排列。根据人们对问题的回答，可以设计一个指标来描述研究对象对共产主义者公民权的观点。显然，认为应当允许共产主义者从事所有这些职业的人比不允许他们从事这些职业的人维护公民权的程度更高；认为应当允许共产主义者从事其中三种职业的人比只允许他们从事一种或两种的人更加维护共产主义者的公民权。于是，指标就提供了这一变量的定序量度。

假定问卷中的职业是这样排列的：（1）挖土工，（2）中学教师，（3）美国总统。对大多数人来说，这三个职业属于趋强结构。一个同意共产主义者当总统的人绝不会反对他当中学教师或挖土工，但一个同意共产主义者当挖土工的人不一定同意他从事其他两种职业。因此在这一研究中，了解某人回答的职业是几种，往往同时就得知他同意的职业是哪几种。本书在这种意义上使用尺度一词。

显而易见，尺度比指标优越，因为它提供了更多的信息；但指标比尺度使用得更普遍，原因是尺度的设计往往更困难，有时甚至不可能设计出来。

2 指标设计

（1）内容的选择

组合指标是用来度量变量的。指标的内容选择的首要标准是逻辑上的正确性。例如你想度量政治保守主义，你选择的每项内容都应与保守主义有关。研究对象的党派身份，赞成还是反对著名保守党人物的观点等等。在选择内容时还要注意差异程度。如果对某项内容来说，任何人都会被确定为保守的（例如对某位任何人都不能接受其观点的极右派人物的看法）这项内容在组成指标时就不宜采用。

（2）内容之间的双变量关系

指标设计的第二步是检验内容之间的双变量关系。例如，研究对象在某项内容上表现出保守倾向，在另一项内容上也应大致属于这种倾向；即使不是如此，在甲内容上表现出保守倾向的研究对象

在乙内容上表现出同一倾向的可能性应当比在甲内容上表现出自由派倾向的研究对象大一些。

显然，如果某项内容与其他内容全不相关则应把它删去；而如果两项内容相关性太强，那么其中的一项也就不必留在组合指标之内了。

试举一例。我们在医学院教授中进行调查，目的在于了解是否重视理论的医生比其他医生处理病人时较少受个人感情的影响。度量“理论型”的指标由三个问题组成：

（1）“作为医学院教授你认为哪种能力对教学更重要：治病能力还是医学研究能力？”

（2）“为进一步增加你的医学知识，你的最终兴趣在于处理病人还是在于对基本医学理论的理解？”

（3）“在医学研究领域你对评价治疗效果的文章更感兴趣还是对探索基本医学理论的文章更感兴趣？”

对每一问题的回答选择第二答案的比选择第一答案的更重理论。表15—1表明三个问题之间的相关关系。三个双变量表囊括了一切由三个问题可能组成的对子。

从表15—1可以看出，三个问题度量的是同一变量：“理论型”。选择“研究者”作为最佳教学

15—1 “理论型”度量标准之间的双变量关系

		最佳教学身份	
		医生	研究者
最终兴趣	处理病人	49%	13%
	理解医学理论	51%	87%
	100%=	(285)	(196)

		阅读兴趣	
		治疗效果	医学理论
最终兴趣	处理病人	68%	30%
	理解医学理论	32%	70%
	100%=	(132)	(349)

		阅读兴趣	
		治疗效果	医学理论
最佳教学身份	医生	85%	64%
	研究者	15%	36%
	100%=	(132)	(349)

身份的教授比选择“医生”的教授更倾向于把自己的最终兴趣确定为“理解医学理论”。数据表明87%选择“研究者”的选择了“理解医学理论”，而只有51%选择“医生”的选择这一答案。这两项内容的强度可概括为“差异36%”。同样，阅读兴趣和最终兴趣之间关系的强度可概括为“差异38%”（70%：32%），阅读兴趣与最佳教学身份之间关系的强度可概括为“差异21%”（36%:15%）。

最初这三个问题是由于逻辑上的正确性被选中的——它们似乎都与理论型倾向有关。通过对每对问题之间关系的检验，证明它们确实可以成为同一变量的度量尺度。

（3）内容之间的多变量关系

在决定将几项内容组合为某一指标之前，应当检验它们之间的多变量关系。如前所述，两项效果完全相同的内容不应当保留在一个指标中，检验内容之间的多变量关系就是为了避免重复内容。仍用医学院教授的例子，表15—2表现了三项内容之间的关系。

表15—2 “理论型”度量标准之间的三变量关系

		最终兴趣为理解医学理论的比例	
		最佳教学身份	
		医生	研究者
阅读兴趣	治疗效果	27% （68）	58% （12）
	医学理论	58% （219）	89% （130）

表15—2与表15—1不同，研究对象按照最佳教学身份和阅读兴趣被分为四组。括号里的数字是各组人数。（66位教授认为最佳教学身份是医生且阅

读兴趣在关于治疗效果的文章）。每个组的百分数表明最终兴趣在于理解医学理论的人所占的比例。（66位教授中，有27％的人最终兴趣在于理解医学理论）。表的左上角是理论倾向最弱的（实践倾向最强的）；表的右下角是理论倾向最强的。可以预料，右下角一组研究对象最终兴趣在于理解医学理论的比例应最大（89％），左上角的比例应最小（27％）。左下角右上角的人居中（58％）。

表15—3 “理论型”度量标准之间的三变量关系（假设）

		最终兴趣为理解医学理论的比例	
		最佳教学身份	
		医生	研究者
阅读兴趣	治疗效果	51％	87％
	医学理论	51％	87％

从表15—2与表15—3（假设表）的对比中，可以看到检验内容之间多变量关系的重要性。表15—3表明，加上阅读兴趣这项内容之后，最终兴趣和最佳教学身份之间的关系不变。在表的每一行中，“研究者”都比“医生”更倾向于理解医学理论，所以一旦得知某个研究对象认为最佳教学身份是医生还是研究者，就不必再去了解其阅读兴趣。由于阅读兴趣不能为这组指标增加任何新的信息，它就

不必与最佳教学身份这项内容并列在指标里了。

这个例子只有三个问题。在更复杂的多变量表中可能出现四个、五个或更多的变量。这一步骤的目的在于发现那些相关程度高的内容并决定是否有必要把它们全部保留在同一指标之中。

（4）指标的分数评定

当你为某一指标选定内容之后，就该为不同的答案评定分数了，各项内容分数之和即为指标。

在评定分数过程中要做出两个决定。第一，决定指标分组的范围。政治倾向的范围可以从“极端保守”到“极端自由”。要注意平衡指标的全距，使每一点上有足够的个案。第二，决定每个问题的具体分数。一般的作法是每个问题分数相等，偶尔也需要做加权处理。在医学院的例子中，可以给每个问题答案为理论型的1分，非理论型的0分。于是每个研究对象可能得到的分数为0、1、2、3，随“理论型”答案多少增减。

（5）未详数需的处理

使用任何方法都会遇到未详数据的问题（如人们填“不知道”）。处理未详数据的方法有很多种，第一，如果未详数据不多，可以把它们删去。

原则是还有足够的数据供分析用及无系统偏差。第二，可以假定未详数据为某一种答案。如某问卷要求研究对象回答是否参加过某几项活动。有人只填写参加过的活动，没参加过的就空在那里。在这种情况下，可以假定空白处是“没有参加”。第三，仔细分析未详数据有可能发现它们所代表的意义。如在某项关于宗教信仰的研究中，人们对某种信仰填“不了解”实际上是“不相信”。要具体情况具体分析。

其他未详数据处理方法包括：设中间数值为未详数据的数值，如在某指标数值为0、1、2、3时，可设未详数据为2；对连续型变量如年龄可用平均数为未详数据的数值；还可用随机数字为未详数据的数值。假定指标总分为6分，某研究对象只回答了4个问题。如果他对4个问题的回答得了4分，可以给他6分；如果他对4个问题的回答仅得了2分（满分的一半），则可以给他3分（满分6分的一半）。

（6）指标的证实

衡量对某一变量的度量成功与否的方法是对指标的证实。证实组合指标可用下列方法：

〔内部证实〕证实指标首先是内部证实。方法是作一个表，以指标为自变量，各项内容为因变

量，看其变化是否一致。如果某项内容与指标相关系数很低则应删去。

〔外部证实〕再以关于医学院教授的研究为例。“理论型”指标高的研究对象在其他问题上也应显示较强的理论倾向；“理论型”指标低的研究对象在其他问题上应显示较低的理论倾向，如表15—4所示：

表15—4 “理论型”指标的证实

	“理论型指标”			
	低			高
	0	1	2	3
参加医学科学讲座的比例	34	42	46	65
持医学院教授应当成为研究者这一观点的比例	43	60	65	89
持医学院教授应参加研究活动这一观点的比例	0	8	32	66
在上学年实际参加研究工作的比例	6	16	24	39

（7）里克特（Likert）尺度

里克特尺度是现代调查问卷中普遍采用的提问格式。它的基本形式是给出一个陈述，要求调查对象表明他是“强烈赞成”、“赞成”、“反对”、“强烈反对”、还是“未决定”。当然有时词语可略有不同（如不用“赞成”而用“同意”）。

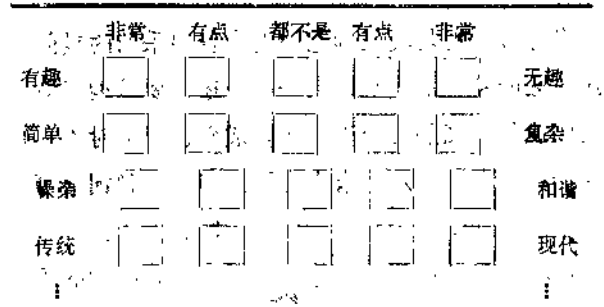
这种格式的特殊价值在于答案具有明确的顺

序，且容易评定分数。有五种答案则评分为0—4或1—5，同时要注意评分的方向（对正面问题给“强烈赞成”者评5分；对反面问题给“强烈反对”者评5分）。

（8）语义差异

语义差异形式要求人们在两个反义词中做选择。例如要评价某种音乐，首先列举几组反义词如“有趣”与“无趣”，“复杂”与“简单”，“和谐”与“噪杂”等等。然后让人们填写如图15—1所示的表格。这里要特别注意，为了避免系统偏差，最好不把看去接近的词放在一侧。如“噪杂”与“传统”被放在左边，“和谐”与“现代”被放在后边，因为那些认为这种音乐噪杂的人往往认为它是现代的。

图15—1　　语义差异



3 尺度设计

好的指标可以在某一变量上给个案排列顺序。指标建立在这样一个假设之上：一位赞成七项保守提案的议员比只赞成四项保守提案的议员更加保守。然而潜在的危险在于，并非所有的内容都同等重要。第一位议员可能赞成七项略微保守的提案，而第二位议员赞成四项极端保守的提案（他可能认为那七项提案保守程度太低而投了反对票）。尺度则可以表达内容的强度。

（1）保迦德斯（Bogardus）社会距离尺度

假定你想了解美国人愿意与阿尔巴尼亚人交往到什么程度，可以问下列问题：

（1）你愿意让阿尔巴尼亚人生活在你的国家吗？

（2）你愿意让阿尔巴尼亚人生活在你的社区吗？

（3）你愿意让阿尔巴尼亚人住在你们那条街吗？

（4）你愿意让阿尔巴尼亚人做你的邻居吗？

（5）你愿意让你的子女与阿尔巴尼亚人结婚吗？

这些问题一个比一个关系更近。这样一组能表达不同强度的问题称为保迦德斯社会距离尺度。

显然，能接受高强度内容的人必定能接受低强度内容。如一个愿意让阿尔巴尼亚人作他的邻居的人绝不会反对他们生活在他的社区和国家，但他不一定愿意让他的孩子同阿尔巴尼亚人结婚。因此，这里就有了一个逻辑结构。

保迦德斯社会距离尺度又是一种很经济的尺度。只要知道某个调查对象接受了几项与阿尔巴尼亚人的关系，也就知道了他接受了哪几项。于是，一个数字可以精确地概括五、六项数据而不会丧失任何信息。

（2）瑟斯滕（Thurstone）尺度

瑟斯滕尺度的基本形式是“间距相等”。把多项内容交给一组裁判，请他们评定每项内容适于做某变量指标的程度。如分数是从1到13，变量是偏见程度，如果裁判认为某项内容很弱则评1分，某项内容很强则评13分，余类推。所有的裁判都完成评定之后，研究者查看某项内容上所有裁判评分的一致程度。评分很不一致的内容就应删去，因为它

们是不明确的。瑟斯滕尺度如今较少被采用，原因在于每位裁判评分的质量取决于其自身的经验与知识，而且变量的重要性也是随时间而改变的。

（3）迦特曼（Guttman）尺度

一种被普遍采用的尺度是迦特曼尺度。象前述两种尺度一样，迦特曼尺度也十分注重度量变量指标的强度。

仍以对医学院教授的研究为例。对迦特曼尺度来说，最重要的是度量理论倾向的指标的强度。最佳教学身份这一指标似乎最强：只有三分之一的人在这个问题上选择“理论型”答案；最终医学兴趣次之：三分之二的人在这个问题上选择了“理论型”答案；阅读兴趣最弱，百分之八十的人选择了“理论型”答案。表15—5是答案模式，正号代表“理论型”答案，负号代表“非理论型”答案。

表的上半部的四种答案模式称为尺度型：这些模式构成尺度结构；表的下半部的答案模式违反了尺度结构，这种混合型的答案不适合于迦特曼尺度。答案符合尺度结构的将得到同样的指标分数和尺度分数。混合型答案则有不同的指标分数与尺度分数。

如第一种混合型答案（－＋－），指标分数

表15—5　　理论倾向尺度

	阅读兴趣	最终兴趣	教学身分	人数
	+	+	+	116
尺度型	+	+	−	127
总计=383	+	−	−	92
	−	−	−	48
	−	+	−	18
混合型	+	−	+	14
总计=44	−	−	+	5
	−	+	+	7

表15—6　　指标和尺度分数

	答案模式	人数	指标分	尺度分	尺度误差
	+++	116	3	3	0
尺度型	++−	127	2	2	0
	+−−	92	1	1	0
	−−−	48	0	0	0
	−+−	18	1	2	18
混合型	+−+	14	2	3	14
	−−+	5	1	0	5
	−++	7	2	3	7

$$\text{再现系数}=1-\frac{\text{误差数}}{\text{人数}\times\text{问题数目}=\text{预测数}}$$

$$=1-\frac{44}{427\times3=1281}$$

$$\approx0.966$$

$$=96.6\%$$

为1,即有一个“理论型”答案，但如果尺度分也给1就会造成他是在强度最低的问题上得分的印象（+－－）。这就错了。因此为混合型答案评尺度分的原则是把误差降到最低限度（见表15—6）。样本中有427个研究对象，对每人在三个问题上做出预测，共为1281个预测。从表15—6中可以看出，运用尺度分数时会产生44个误差，正确预测的比例称为再现系数：再现这些答案的比例。在这个例子中，再现系数为1237/1281，或96.6%。

4 分 类

前述指标和尺度都是单维量度，有时需要做二维或多维量度。例如，你想分别了解报刊在国内政策和外交政策两方面的政治倾向。表15—7为四重分类表。

表15—7　　报刊的政治分类

	外交政策	
国内政策	保守	自由
保守	A	B
自由	C	D

A类报刊是在外交和国内政策上都持保守态度，D类报刊是对内对外都持自由态度，B类和C类在一方面保守，另一方面自由。

分类往往能够帮助我们理解数据，但应当注意最好以分类变量为自变量而不是因变量。

第十六章
阐释模式

1 阐释模式发展史

阐释模式的起源对于科学研究在实践中的应用具有启发性。在第二次世界大战期间，塞缪·斯托芬（Samuel Stouffer）组织和领导了对美军内部影响士气因素的研究。这个研究首先检验了一些公认的模式，例如：

（1）由于军阶的升迁影响士气，升迁率低的部门士气会相对低些。

（2）由于美国南部的种族隔离和歧视，北部的黑人士兵比南部的黑人士兵士气高些。

（3）教育程度越高的人越不愿意当兵。

出人意料的是，研究结果表明，这些公认的模式都是错的：首先，军事警察在军中是升迁率最低的部门；空军是升迁率最高的部分，可是前者比后

者抱怨升迁太慢的人要少。此外，黑人士兵的士气并不因在南部还是在北部而不同。第三，教育程度越低的人越不愿当兵。

原因何在？斯托芬发现答案在于参照系和相对剥夺感。人们并非根据绝对的、客观的标准评价自己的地位，而是与周围的人相比，即与其参照系相比，当他们比不上这些人时才会产生相对剥夺感。

运用参照系和相对剥夺感理论，斯托芬发现了前述与一般看法相悖的现象的原因。在升迁问题上他认为士兵评价升迁制度公平与否是基于个人的经验。在军事警察部门，升迁又少又慢。这个部门的士兵很少发现水平低于他的人比他提升得快，而在空军中，相对快速的提升使很多人发现有些水平不如自己的人得到提升。因此，军事警察比空军士兵对升迁制度的不满较少。

类似的分析方法也可解释黑人士兵的状况。在南方的黑人士兵不是把自己的状况同北方的黑人士兵比，而是同他们周围的黑人公民比。美国南部种族歧视很严重，但参军可以同一般的文化规范隔离开来，因而黑人士兵比一般黑人公民境况要好。在北方，许多黑人能得到更好的工作，参军并不能提高其社会地位。

最后，斯托芬用参照系概念解释了教育程度低

的人反而不愿当兵的原因：教育程度低的人多为从事工农业生产的半熟练工和农民。在战争期间他们的工作是国家急需的因而可以缓服兵役，而且他们多有在这些缓征行业工作的朋友。当同朋友相比时，教育程度低的人被征入伍就会觉得格外吃亏。

斯托芬的逻辑阐述为阐释模式的发展奠定了基础。其主旨在于通过其他变量来解释两个变量之间的关系。阐释模式的正式形成则归功于保尔·拉扎斯费德（Paul Lazarsfeld）及其助手的工作。他们从斯托芬关于教育程度与参军意愿的正比关系数据着手（见表16—1）制作了一组假想的数据，解释教育程度与朋友得到缓服兵役许可之间的关系。从表16—2可以看出，只有19%教育程度高的人说他们有朋友缓征，而有79%教育程度低的人有这种情况。

斯托芬的理论认为，有朋友被批准缓征的人比

表16—1　斯托芬关于教育程度与参军意愿之间关系的数据

	教育程度高	教育程度低
不要求缓征	88%	70%
要求缓征	12%	30%
	100%	100%
	(1,761)	(1,896)

表16—2　教育程度与朋友缓征之间的关系（假设）

	教育程度高	教育程度低
有朋友缓征	19%	79%
无朋友缓征	81%	21%
	100%	100%
	（1,761）	（1,676）

没有这种情况的人更不愿参军（见表16—3）

表16—3　朋友缓征与本人参军意愿之间的关系（假设）

	有朋友缓征	无朋友缓征
不要求缓征	63%	94%
要求缓征	37%	6%
	100%	100%
	（1,819）	（1,618）

表16—2和表16—3试图证明斯托芬的理论。第一，教育程度低的人比教育程度高的人有更多的朋友被批准缓征；第二，有朋友缓征的人更倾向于要求缓征。斯托芬曾指出这两点是教育程度与参军意愿之间关系的解释。拉扎斯费德用其假设数据注解了斯托芬的理论（见表16—4）。

从表16—4可以看出，教育程度对参军意愿没有影响：有朋友缓征的人虽教育程度不同，但不要求缓征的比例相同（63%）；在无朋友缓征的人中也出现同样情况（94%和95%）。参军意愿受到是

表16—4　通过是否有朋友缓征解释教育程度与参军意愿之间的关系（假设）

	有朋友缓征		无朋友缓征	
	教育程度高	教育程度低	教育程度高	教育程度低
不要求缓征	63%	63%	94%	95%
要求缓征	37%	37%	6%	5%
	100%	100%	100%	100%
	（335）	（1,484）	（1,426）	（392）

否有朋友缓征的强烈影响（63%:94%）。是否有朋友缓征这一变量并未推翻教育程度与参军意愿的关系，只是解释了这种关系之所以产生的原因。这就是阐释模式与多变量分析的精华所在。

通过其他变量的影响理解两变量之间关系性质的主要方法是把样本按照控制变量（又称检定变量）分组。如是否有朋友缓征这一变量就是前例中的控制变量，因而我们把样本按照有朋友缓征和无朋友缓征分为两组。在这两组中分别描述两个初始变量之间关系的表格称为分表，从中发现的关系称为“部分关系”。最后用部分关系同整个样本中的初始关系加以比较。

2 阐释模式

首先必须搞清控制变量的作用是发生在另外两个变量之前还是在它们之间。如果控制变量在中间，如前例的教育程度、朋友缓征和参军意愿那样，分析将以图16—1中的模式为基础进行。这一多变量关系的逻辑是：自变量（教育程度）影响控制变量（是否有朋友缓征），而后者又影响了因变量（参军意愿）。

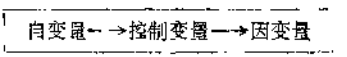


图16—1　位于中间的控制变量

如果控制变量是自变量与因变量关系的前提，模式则不同（见图6—2）。在这种情况下，控制变量对“自变量”“因变量”都有影响。严格地说，它们不能再被称为自变量和因变量了。实际上是一个自变量（控制变量）和两个因变量。虽然表面上“自变量”和“因变量”相关，实际上却并无因果关系。它们表面上的相关是由于各自与控制变量有关而造成的。

表16—5是理解阐释分析方法的指南。表的右边表明控制变量是作为前提还是位于中间；表的左

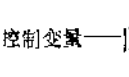


图16—2　作为前提的控制变量

边表明部分关系与初始关系比较后的结果。表主体中所用术语下面将一一讨论。

表16—5　　解释模式

部分关系与初始关系之比较	控制变量	
	作为前提	位于中间
关系相同	重述	
关系减少或消失	解释	注释
关系分裂*	区分	

*部分相同或增加，部分减少或消失。

（1）重　述

不管控制变量是作为前提还是位于中间，只要部分关系与初始关系基本相同，我们就用重述一词来表达这种结果，意为虽然增加了第三个变量，前两个变量之间的关系没有改变。如前例，假如教育程度在有朋友缓征与无朋友缓征两组对参军意愿影响相同，则初始关系被重复了一遍。

（2）解　释

解释一词用于描述伪关系——由于引进控制变

量，初始关系被解释了。此类关系中的两个必要条件是：控制变量必须在自变量因变量之前；部分关系必须为零或比初始关系显著减少。举一个简单的例子。在救火车的数量和火灾造成的损失之间有一正比关系：火灾中救火车出动得越多，火灾损失越大。有人可能因此认为救火车数量与火灾损失之间有因果关系。然而，作为前提的控制变量——火灾的规模——解释了这种初始关系。大火灾比小火灾造成的损失大，同时需要更多的救火车。如果把大火灾和小火灾分成两组来看，初始关系就消失了（甚至会出现反比关系——救火车出动越多；火灾损失越小）。

（3）注　释

注释与解释很相近，唯一的不同在于控制变量位于自变量与因变量之间。前面关于参军意愿的研究就是一个出色的例子。教育程度对参军意愿的影响并未消失，这种关系依然存在。归根结底还是教育程度造成不同的参军意愿，中间变量即是否有朋友缓征只是帮助说明即注释了此种关系发生的过程。

（4）区　分

有时几种部分关系明显不同。例如有的部分关

系与初始关系非常接近，有的部分关系中初始关系却消失了，这种情况称为区分。在一个关于教徒来源的研究中，葛洛克（Glock）及其助手发现，教徒的数量与阶级地位成反比。在一个教区中，地位越低的阶层教徒数量越多。葛洛克认为这是因为阶级地位低（由收入和教育程度来度量）的人在世俗社会中得到的满足感比阶级地位高的人要少。如果这一假设成立，那么能够在世俗社会中获得满足感的人们当中，阶级地位与教徒数量之间的初始关系应当消失。他以在世俗组织中拥有办公室作为在世俗社会中获得满足感的指标，将其作为控制变量。研究结果表明，在那些拥有办公室的人当中，阶级地位与教徒数量之间的关系确实消失了；而在没有办公室的人当中，这种关系仍旧存在。（见表16—6）阶级地位与教徒数量之间的初始关系被限制在一种特殊的条件下：只限于那些在世俗社会得不到

表16—6　阶级地位对教徒数量的影响（以是否在世俗组织中拥有办公室为控制变量）

	不同阶级教徒比例				
	低				高
	0	1	2	3	4
有办公室	46%	53%	48%	46%	46%
无办公室	62%	55%	47%	46%	40%

满足的人。

（5）模式的改进

莫瑞斯·罗森伯格（Morris Rosenberg）对阐释模式的改进作过出色的表达。首先，假定在两个变量中存在着某种初始关系。他甚至主张，即使在初始关系为零的情况下也可以使用阐释模式。他引用了一个关于参加工会时间长短与对吸收犹太人为工会工作人员之间关系的研究（见表16—7）从最初的分析看，似乎二者不相关，其实是年龄压抑了

表16—7　　压抑变量一例

Ⅰ．表面上对犹太人态度与参加工会时间长短无关

	入会少于四年	入会四年以上
同意犹太人成为工会工作人员	49.2%	50.4%
	（126）	（256）

Ⅱ．在各个年龄组，入会时间与接受犹太人为工会工作人员的意愿成正比

年龄与入会时间					
29岁以下		30—49岁		50岁以上	
入会时间					
少于四年	四年以上	少于四年	四年以上	少于四年	四年以上
同意犹太人成为工会工作人员	56.4%（78）	62.7%（51）	37.1%（55）	48.3%（116）	38.4%（13）	56.1%（89）

二者之间的关系，因而年龄在这里又称为压抑变量，它掩盖了入会时间与对犹太人态度之间的关系。

其次，分析部分关系是否与初始关系相同或比它微弱。第三．分析部分关系是否比初始关系更强或与其方向相反，如，初始关系为反比关系，而部分关系为正比关系。

罗森伯格还用一个假想的例子引入歪曲变量这一概念：在初始关系中，工人阶级似乎比中产阶级更支持公民权运动，但后来发现种族是歪曲变量，它歪曲了阶级和支持公民权程度之间的真实关系。

表16—8　　歪曲变量一例（假设）

Ⅰ．工人阶级比中产阶级更支持公民权运动

支持公民权程度	中产阶级	工人阶级
高	37%	45%
低	63%	55%
	100%（120）	100（120）

Ⅱ．控制种族变量后发现中产阶级比工人阶级更支持公民权运动

	社会阶级			
	黑人		白人	
支持公民权程度	中产阶级	工人阶级	中产阶级	工人阶级
高	70%	50%	30%	20%
低	30%	50%	70%	80%
	100%(20)	100%(100)	100%(100)	100%(20)

黑人比白人更支持公民权运动，但黑人多为工人阶级，很少中产阶级。中产阶级的黑人比工人阶级黑人更支持公民权运动，白人中也是如此。如果控制种族变量，将会得到中产阶级比工人阶级更支持公民权运动的结论。

第十七章
社会统计学

本章将讨论统计学的两个基本类型：描述与推论。描述统计学是描述数据的媒介；推论统计学帮助我们从观察中抽取结论，一般来说是从样本中抽取关于总体的结论。

1 描述统计学

（1）数据的简化

表17—1是一个数据矩阵，每列代表一个变量。如，第三列代表性别，1为男，2为女。在第十四章曾经讲过，原始数据矩阵的优点在于囊括一切信息，其缺点是并未有效地将数据分布表达出来。假如你有200个变量，2000个研究对象，数据将多达40万个。很难想象从如此浩繁的数据中直接发现什

么有意义的模式。

表17—1　　典型的原始数据矩阵

	变量1	变量2	变量3	变量4	变量5……	变量n
个案1	2	5	1	4	3 ……	6
个案2	3	3	1	1	3 ……	1
个案3	1	3	2	3	2 ……	2
个案4	2	1	2	2	5 ……	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
个案n	1	3	1	2	4 ……	4

描述统计学提供了简化数据矩阵的方法，对原始数据做出概括，以便做进一步的理解和阐释。在第十四章已经讨论过单变量数据的概括方法：频数分布与频率分布、分组、平均数、中位数、众数、全距、标准差等等。描述统计学的基本目标在于效率：以最简单的概括形式最大限度地容纳信息。

（2）相关的度量

数据矩阵也可用来描述两变量之间的相关关系。表17—2是双变量频数分布表，为确定变量X和Y之间关系的性质提供了一切必要的信息。X为列，Y为行。表内数字表明取值的特殊模式。如：35个研究对象具有X_1Y_1的模式，43个具有X_3Y_7的模式。描述统计学有多种方法对数据矩阵进行概括，哪种方

法最佳视变量性质而定。

以下讨论的多种相关度量方法全都以消减误差比为基本原则。假定我们要推测研究对象对某一问题的回答是肯定的还是否定的。在第一种情况下，假定我们知道样本答案的分布，60%为肯定，40%为否定，对每个研究对象的答案都推测为“肯定”（频率最高的取值）就能最大限度地消减误差。在第二种情况下，假定我们还知道这个变量与另一变量(如性别)之间的关系，在了解某人性别的情况下推测其答案是肯定还是否定。如果这两个变量是相关的，推测误差就会减少。了解了两变量之间的相关程度就可以计算出消减误差比：相关程度越高，消减误差比越大。

表17—2　　用数据矩阵描述变量之间的相关关系

	变量X					
	X_1	X_2	X_3	X_4	X_5	X_6
Y_1	35	27	26	12	15	7
Y_2	38	48	38	22	35	13
Y_3	52	41	75	64	46	22
Y_4	28	45	63	80	79	45
Y_5	20	35	58	90	103	87
Y_6	23	12	76	80	99	165
Y_7	5	8	18	60	75	182

消减误差比随不同的度量层次而略有不同。下面按定名、定序及定距等不同度量层次分别介绍最适宜的方法。当然这三种度量方法仅是大量度量方法中的一小部分。

〔定名变量〕如果两个变量都是定名变量（如性别、宗教信仰、种族等），可用λ来度量。λ代表已知变量值对未知变量值的推测力。表 17—3是一

表17—3　　性别与就业状况（假设）

	男	女	总计
就业	900	200	1100
失业	100	800	900
总计	1000	1000	2000

组性别与就业状况的假设数据，其中1100人为就业者，900人为失业者。如果要求你仅仅根据这两个数字来推测研究对象的就业状况，你应当全部推测为“就业”，因为这种推测比全部推测为“失业”误差少些。即使如此，在2000个推测中仍会产生900个错误。

如果你掌握了表17—3的信息，且了解每个研究对象的性别，则推测方法应稍加改变：对每位男性推测为“就业”，女性推测为“失业”。于是误差减少到300个——100个失业男性和200个就业女

性，推测误差比不知性别时减少600个。

λ是消减误差与全部误差之比。本例中λ为0.67；600（消减误差）除以900（仅据就业状况推测所产生的全部误差）。λ以此度量性别与就业状况之间的统计相关。如果性别与就业状况不相关，情况应当是男性与女性的就业状况相同。在此种情况下，了解了研究对象的性别并不能消减推测其就业状况的误差，λ于是为零。如果男性全部就业，女性全部失业，了解了性别就能消减推测其就业状况的全部误差，于是λ为1，表示完全的统计相关。

〔定序变量〕如果两个变量都是定序变量（如社会阶级、宗教信仰程度、异化程度等）则可用γ来度量。与λ相似，γ也代表已知变量值对未知变量值的推测力。区别在于，γ是通过研究对象在某一变量上的大小顺序去推测其在另一变量上的大小顺序。假定宗教信仰程度与政治保守程度成正比关系，已知甲比乙宗教信仰程度高即可做出甲比乙政治上更加保守的推测。

表17—4是一组社会阶级地位与偏见程度的假设数据。而变量关系的一般性质是当社会阶级地位增高时，偏见程度降低。二者呈反比关系。

γ是由以下两个数值计算出来的：（1）同序对数量，（2）异序对数量。计算同序对做量的方法

表17—4　　社会阶级地位与偏见程度（假设）

偏见程度 \ 社会阶级地位	下层	中层	上层
低	200	400	700
中	500	900	400
高	800	300	100

是用表内每个频数分别乘以其右下方全部频数之和，再把这些结果相加。如表17—4，同序对数量为200（900＋300＋400＋100）＋500（300＋100）＋400（400＋100）＋900（100）＝830,000。计算异序对数量的方法是用表内每个频数分别乘以其左下方全部频数之和，再把这些结果相加，即700（500＋800＋900＋300）＋400（800＋300）＋400（500＋800）＋900（800）＝3,430,000。γ公式为：

$$\gamma = \frac{\text{同}-\text{异}}{\text{同}+\text{异}} \quad \text{即} \quad \frac{830,000-3,430,000}{830,000+3,430,000} = -0.61$$

答案中负号表明二变量间存在反比关系。γ数值则表明异序对的数量比同序对多出它们总和的61%，γ值在－1与＋1之间，除表明二变量相关程度还表明了相关的方向。

〔定距或定比变量〕如果两个变量都是定距或定比变量（如年龄、收入、平均分数等）可用皮尔

逊（Pearson）相关系数r。从某种程度上可以说，r表明用已知变量值推测未知变量值的准确程度。

在推测时，对定名变量用众数，对定距或定比变量则需要用平均数。这样做虽然不能保证推测全部正确，但可把误差消减到最低限度。误差将由实际数值与平均数之差的平方和来度量，又称全部误差。为了提高推测的精确度，应作回归线（见第十八章）利用回归方程即可由一个变量值对另一变量值作出估计。方程的一般形式为$Y'=a+b(x)$，其中a和b是常数，X是某变量的一个数值，Y'是对另一变量对应值的估计。Y与Y'之间的差异即Y的实际数值与估计值之差的平方和是无解误差，又称剩余向量，它是用X值估计Y值时的误差。无解误差与有解误差之和为全部误差。有部误差与全部误差之比为消减误差比。前例$r=0.7$，$r^2=0.49$，这个数字表明大约一半误差是可以解释的。

〔混合变量〕研究中常常会遇到相关变量类型不同的情况，如一个是定序变量，另一个是定名变量。每种不同的组合都有其最适宜的度量方法。限于篇幅不详述了。

2 推论统计学

许多社会科学研究项目都涉及对样本数据的分析。研究者研究样本并非为了描述样本本身，而是为了推论总体。

（1）单变量推论

第十四章曾介绍过描述单变量数据的方法。本节介绍两种单变量推论方法：比率和平均数。

如果样本中50%的人说他们在去年得过感冒，那么50%这个数值就是对总体中患感冒人数比例的最佳估计。当然在总体中不会是精确的50%，但只要抽样设计遵循了随机抽样的原则，就可以对样本推论总体时可能的误差范围作出估计。

标准差对抽样误差的估计是极其重要的。总体数值落在样本数值正负一个标准差范围内的概率是68%；落在正负两个标准差范围内的概率是95%；落在正负三个标准差范围内的概率是99.9%。

对抽样误差的估计包括两项基本内容：置信水平（如95%）和置信域（如±2.5%）。如果在容量为1,600的样本中有50%的人说他们得过感冒，那么

总体数值落在47.5%与52.5%之间的概率是95%。

推论的几个重要前提是：第一，样本必须是从待推论的总体中抽取的。从电话簿上抽取的样本不能成为关于一个城市总体的统计推论的基础。第二，抽样应当是随机的。第三，统计推论只适用于抽样误差，不适用于非抽样误差。总体中47.5%—52.5%的人将报告得过感冒的概率是95%，但他们是否实际上得了感冒就不得而知了。由于非抽样误差往往比抽样误差还大，必须在推论时考虑到这一因素。

（2）统计显著性的检验

判定两变量相关的显著程度的方法之一是显著水平的参数检定。如前所述，样本统计值是总体参数值的最佳估计，但二者并不完全相同，因此我们报告参数值落在一定范围（置信域）内的概率。下面用图解来解释统计显著性检验的逻辑。它包括以下几个因素：1. 总体的两个变量不相关；2. 根据古典概率抽样程序抽取样本；3. 样本分布是两个变量的二维分布。

图17—1是一个数量为900的总体，每个点子是总体的一个成员，边线是总体的概念边界。在概念的空间，个体是均匀分布的。

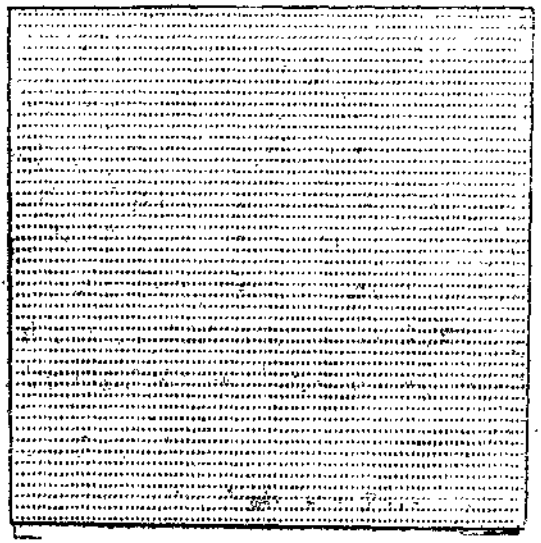


图17—1　包括900个个案的总体（假设）

图17—2是变量X的两个取值的总体分布。在概念的空间，个体仍是均匀分布的。

图17—3加上了另一个变量（Y）。总体中二维分布的个案表明两个度量是不相关的，即一个个案是X_1还是X_2对其是Y_1还是Y_2没有影响。

图17—4是从两变量不相关的总体中抽取的容量为400的随机样本。在总体中央的正方形表示样本对总体的代表性。这一图形的形状与总体差异越大则代表性越差。图17—4总体中央的正方形表示样

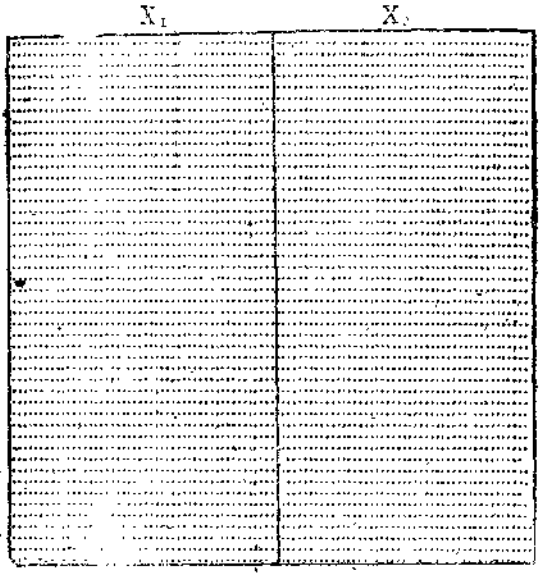


图17—2 在概念空间的总体分布，变量的均匀分布

本对总体具有完美的代表性，其中X和Y是不相关的。

图17—5是更接近实际情况的抽样结果，其中样本没有完美地代表总体。随机抽样的基本逻辑包括某种程度的抽样误差。

图17—6中，抽样结果与XY不相关的假设相悖。在取值为X_1的个案中，75%为Y_1；而取值X_2的个案中只有25%为Y_1。样本数据表明X和Y之间存在着强烈的正相关关系。对于两变量不相关的假

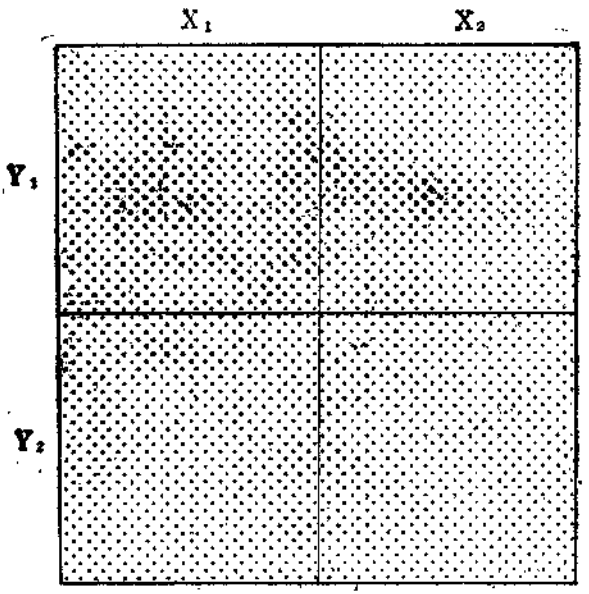


图17—3　在概念空间的总体分布：两变量完全不相关

设来说，这一样本基本上不代表总体。

图17—7表明样本数据否定了总体中XY 两变量不相关的假设，即总体中两变量象样本中一样是有相关关系的。这就涉及了统计显著性检验的基本逻辑：一旦发现总体中两变量不相关这一假设与样本数据分布不一致时，有两种方法来解释这种差异：（1）归因于样本缺乏代表性；（2）否定变量之间不相关的假设。

总体			样本		
	X_1	X_2		X_1	X_2
Y_1	50%	50%	Y_1	50%	50%
Y_2	50%	50%	Y_2	50%	50%
总计	100%	100%	总计	100%	100%

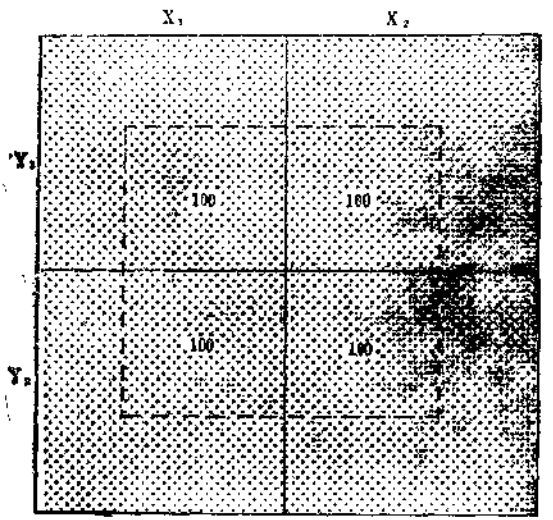


图17—4　从两变量不相关的总体中抽取的具有完美代表性的样本

样本数据之间关系的统计显著程变常常用概率来表达。显著水平为0.05($P \leq 0.05$) 意为总体参数

总体			样本		
	X_1	X_2		X_1	X_2
Y_1	50%	50%	Y_1	50.5%	49.5%
Y_2	50%	50%	Y_2	49.5%	50.5%
总计	100%	100%	总计	100%	100%

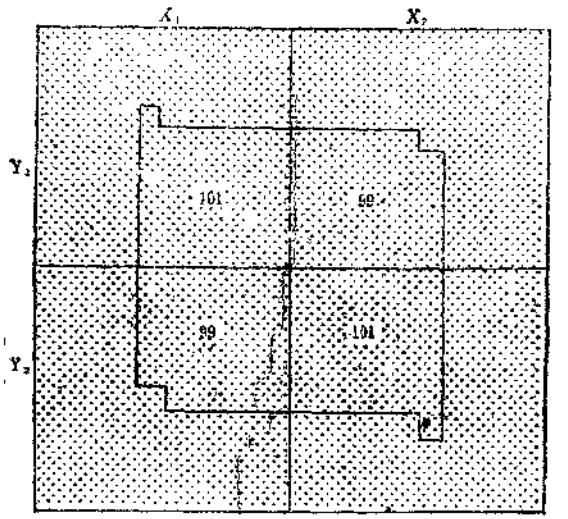


图17—5　从两变量不相关的总体中抽取的代表性不完美的样本

值与样本观察值具有同等相关关系的概率达95%，抽样误差不超过 5%。常用的显著水平有三个：

总体			样本		
	X_1	X_2		X_1	X_2
Y_1	50%	50%	Y_1	75%	25%
Y_2	50%	50%	Y_2	25%	75%
总计	100%	100%	总计	100%	100%

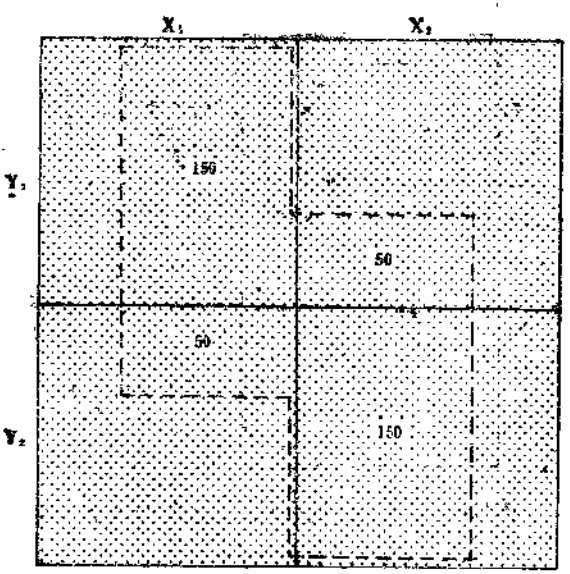


图17—6　从两变量不相关的总体中抽取的基本不代表总体的样本

总体			样本		
	X_1	X_2		X_1	X_2
Y_1	75%	25%	Y_1	75%	25%
Y_2	25%	75%	Y_2	25%	75%
总计	100%	100%	总计	100%	100%

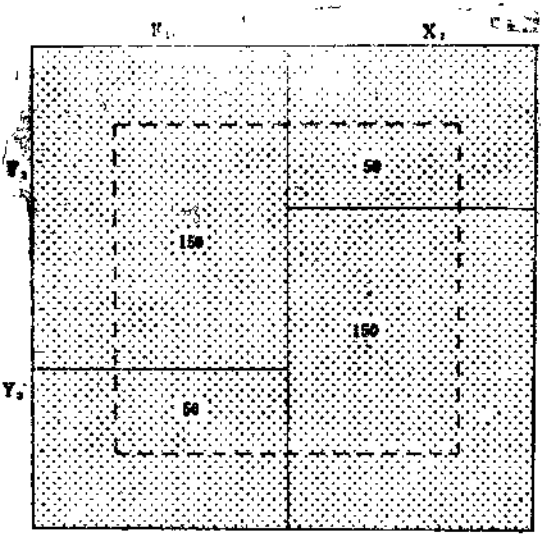


图17－7　从两变量强相关的总体中抽取的具有完美代表性的样本

0.05、0.01和0.001，意为抽样误差概率为5/100、1/100和1/1000。

显著性检验一般有两种作法。一种是在检验之前规定显著水平。如计算出的相关程度达到这一水平则确认两变量相关。另一种作法是报告每一相关系数的显著水平，如0.023，即抽样误差的概率为23/1000。

卡方（X^2）检验是社会科学中常用的显著性检验方法。它首先以原假设（又称零假设，H_0）为基础，假设在总体中两变量不相关。然后计算在假设两变量不相关情况下的期望值，并由期望值与样本观察值的差计算卡方值，确定仅由抽样误差造成的差异的概率。

假定我们想研究上教堂作礼拜与教徒性别之间的关系。首先随机抽取一个包括100名教徒的样本，其中40名男性，60名女性。样本中70%的人报告说上个星期做了礼拜，其余30%未做礼拜。如果性别与做礼拜不相关，男女各应有70%做了礼拜，30%未做礼拜。表17—5的第一部分是期望值，即28位男性和42位女性做了礼拜，12位男性和18位女性未做礼拜。表17—5的第二部分是样本观察值。样本中的40位男性，20位做了礼拜，20位未做；女性中50位做了礼拜，10位未做。

X^2的计算步骤如下：(1)用观察值减去期望值，(2)平方，(3)除以期望值，然后将各项结果相加

表17—5　　卡方检验（假设）

Ⅰ.期望值	男	女	合计
做过礼拜	28	42	70
未做礼拜	12	18	30
合　计	40	60	100

Ⅱ.观察值	男	女	合计
做过礼拜	20	50	70
未做礼拜	20	10	30
合　计	40	60	100

Ⅲ.(观察值--期望值)^2÷期望值

	男	女	
做过礼拜	2.29	1.52	$X^2=12.70$
未做礼拜	5.33	3.56	$P<0.001$

得到X^2值为12.70。确定X^2值是否显著，需要运用X^2检验临界值表，本例的相关显著性在0.001水平上。

第十八章
复杂分析方法简介

本章介绍七种较为复杂的分析方法：回归分析、时间系列分析、路径分析、因素分析、最小间隔分析、方差分析和对数线性模式，为进一步学习这些方法提供参考。

1 回归分析

我们已经提到过描述两变量相关关系的一般公式 $Y=f(X)$，读作Y是X的函数，意为Y值可以由X值的变化来解释。换言之，X导致Y，X值决定Y值。回归分析包括下列几种形式：

（1）线性回归

当两变量间存在完美的线性相关时，回归模式

最明显。图18—1是假设的变量X和Y的分布图。它表明在四个个案中，X值和Y值完全相同。两变量关系的方程为Y = X，这就是一个回归方程。四个点都落在一条直线上，在这些点上做一条直线就是回归线。

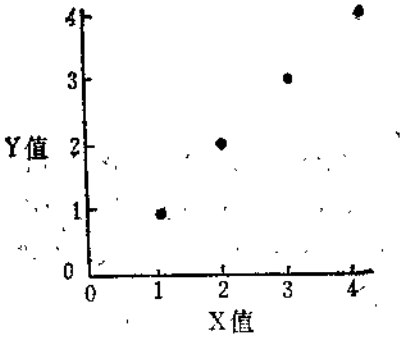


图18—1　X值和Y值的分布图

线性回归模式具有重要的描述功能，它提供变量X和Y之间关系的图形，回归方程是对这种关系的充分概括。回归模式还具有推论价值。根据回归方程对两变量一般关系的描述，可推测具体的数值。如前例，已知某个案X值为3.5，可推知Y值也是3.5。

当然在实际情况中很难遇到如此完美的线性关系。图18—2是一个更实际的例子——中小城市人口与犯罪率之间的关系（假设）。分布图中的每个

点子是一个城市，点子的座标代表其人口及犯罪率。Y值（犯罪率）与X值（人口）相对应，当X值增大时Y值也增大。

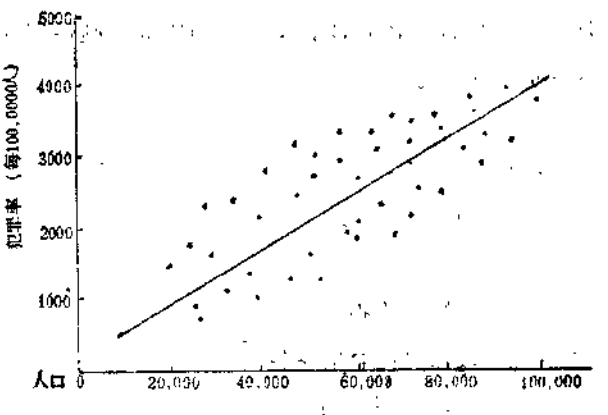


图18—2　两变量值的分布图及回归线（假设）

在图18—2上不可能做出一条通过每个点的直线，但可做出一条线近似地代表每个点，其方程式为$Y = a + bX$，X和y是两变量的值。在这个方程式中，当X为0时a等于Y值，b是斜率。如a值和b值已知，即可计算出与每个X值相对应的Y的估计值。在本例中，如果已知某城市的人口数，即可推算出其犯罪率。

（2）多元回归

社会科学家常常会遇到这种情况：一个因变量同时受几个自变量的影响。多元回归分析方法适用于这种情况。贝沃里·叶格（Beverly Yerg）曾用这一方法做过对体育课教学效果的研究。她使用的多元回归方程是：

$$F = b_0 + b_1 I + b_2 X_1 + b_3 X_2 + b_4 X_3 + b_5 X_4 + e$$

F = 学生期末总分

I = 学生初始分数

X_1 = 对练习的指导

X_2 = 教师对内容的掌握

X_3 = 对特殊问题的反馈

X_4 = 简单明确的示范动作

b = 回归系数

e = 剩余因素

在这一线性回归方程中有多个X和b，剩余因素e是X不能对Y作出解释的部分。叶格计算出复相关系数作为推测学生期末总分的指标，一般用大写R来表示。本例中R = 0.877，意为总分的77%（R^2）可由诸自变量来解释。

（3）部分回归

我们曾讨论过如何在研究两变量关系时控制第三个变量。如在研究教育程度对偏见程度的影响时控制年龄变量，具体作法是分别在各个年龄组内计算教育程度对偏见程度的影响，部分回归建立在同样的逻辑基础上。在计算变量间关系对控制第三变量，然后将其结果与未控制第三变量的结果加以比较。

（4）曲线回归

曲线也可由方程表达，如方程式 $X^2 + Y^2 = 25$ 是一个半径为5的圆。对某些现象的描述用曲线回归比直线回归更合适。按一个复杂方程做出的一条古怪曲线也许能恰好把全部数据联结起来，但并不能据此对新的观察值做出正确的推测。因此，曲线回归虽然为线性回归分析增加了一种选择，但并不能解决一切问题。没有一种方法是能够解决一切问题的。

2 时间系列分析

上述各种形式的回归分析方法亦常用于处理时间系列数据，以说明一个或多个变量在一段时间内的变化。例如美国的犯罪率就是逐年上升的。犯罪率的时间系列分析能够以回归形式表达长期趋势并解释这一趋势，如人口增长和通货膨胀的趋势及其对犯罪率的影响。

图18—3是某城市的失窃率(假设)。图内每个点代表当年报警的失窃数字。假定失窃是经济状况特别是失业率的函数——失业时人容易沦为窃贼，于

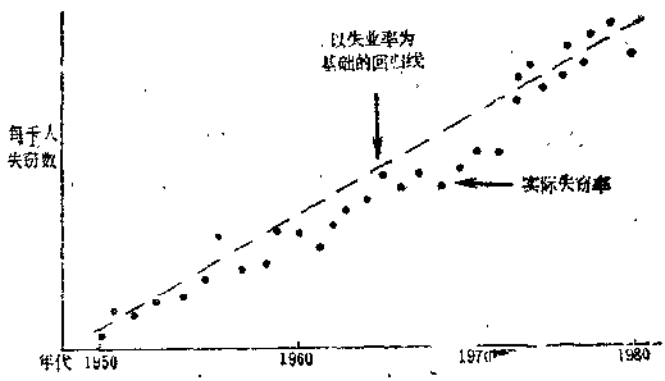


图18—3 随时间变化的失窃率(假设)

是我们设计一个回归方程来表达失窃率与失业率之间的关系。图中虚线即是假设的回归线。

情况往往并不如此简单。如人们失业之后并非立即开始盗窃而是先花光积蓄，向朋友借债，等待就业机会，最后实在无生路才走上盗窃的道路。滞后回归分析就是专门用来处理这种复杂情况的。例如，可以前几年的失业率为基础做回归线，预测某一年的失窃率。回归方程设计的成败得失应由因变量的实际观察值来判断。

3 路径分析

路径分析是理解变量间关系的因果模式。它也是以回归分析为基础，但能比其他方法更形象地表达出诸变量间的关系。路径分析假定某变量影响另一变量，因此必须区分自变量和因变量。除画出变量关系网之外，还可计算出这些关系之强弱，表达关系强弱的数字称为路径系数，是用回归分析的方法计算出来的。

试举一例。图18-4是关于基督徒中宗教信仰程度造成反犹主义的分析。从左向右图中的变量为（1）基督教正统观念或研究对象接受传统的关于

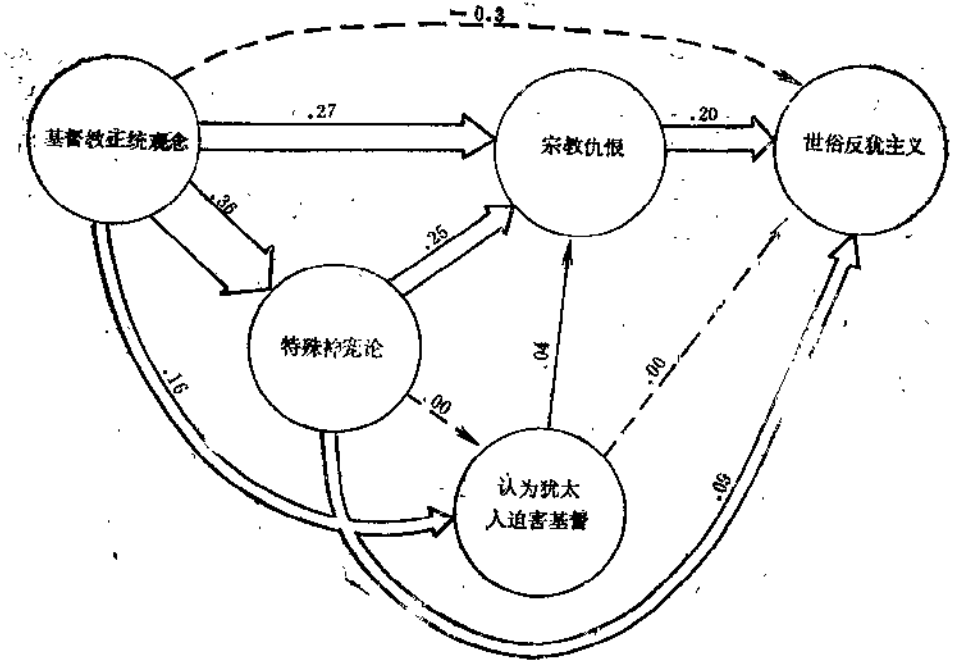


图18—4　反犹主义的宗教根源

上帝、耶稣基督、圣经奇迹等信条的程度；（2）特殊神宠论，即相信自己所信的宗教是“唯一真理”的程度；（3）接受犹太人在历史上迫害过基督这种说法的程度；（4）对现代犹太人的宗教仇恨程度，如相信上帝正在惩罚犹太人，犹太人除非改信基督教必受诅咒等等；（5）世俗的反犹主义，如认为犹太人多为奸商、对国家不忠诚等等。

研究开始时，研究者假设这五个变量之间的关系是：正统观念导致特殊神宠论，再导致犹太人在历史上迫害过基督的说法，再导致宗教仇恨，最后导致世俗的反犹主义。但从路径图上却发现了不同的因果关系。如认为犹太人在历史上迫害过基督这一点对整个因果过程无大影响，而反犹主义是更直接地由正统观念和宗教仇恨造成的。

4 因素分析

因素分析可用来在几个变量之间发现关系模式，其主要方法是从变量中概括出几个人为设计的尺度（因素）然后用计算机进行处理。假定某数据档案里有关于研究对象偏见程度的几个指标。每个指标都可代表偏见程度的一部分，但没有一个指标

可以单独成为偏见程度的代表，且这几个指标互相高度相关。在因素分析中，研究者人为地制造一个总尺度，它与各个分指标高度相关。每个研究对象在这个人为尺度上有一个数值，这个数值即是其在各个分指标上的最佳估计值。

因素分析的结果包括从变量关系中概括出来的几个因素同各个变量之间的相关系数，又称因素负荷。对某一特定因素来说，读者可以很容易看出哪些变量是其高度负荷，哪些变量是其轻微负荷。

因素分析的缺点在于：第一，因素没有实体的意义。例如，对于某个特定因素来说，偏见程度和宗教信仰程度同它正相关，而教育程度却同它反相关。缺乏经验的研究者在这种情况下会感到束手无策，第二，因素分析常常受到来自一般哲学方面的批评。它们认为，某项陈述如能成立必须具有可证伪性，否则不是同义反复就是毫无意义。而在因素分析中，输入任何数据都能得到某种模式。这一事实在评价因素分析的结果时必须考虑进去。对因素的概括并非总是有意义的。

5 最小间隔分析

最小间隔分析是一种较新但发展潜力极大的分析方法。它建立在变量间相关系数的基础上，任何相关系数的度量方法都可用于最小间隔分析，下面仅以皮尔逊（Pearson）的r为例。

让我们从一个描述三变量A、B和C之间关系的简单矩阵开始（见表18—1）。如果两变量相关系数高就将其摆得间隔小些；相关系数低就将其摆得间隔大些。于是从表18—1得到：

A　　B　　C

表18—1　　简单相关矩阵

	A	B	C
A	x	0.8	0.2
B		x	0.5
C			x

由于A与B相关系数最高，其间隔最小；其次是B和C；图中离得最远的是A和C，二者相关系数最低。

在矩阵中加上变量D（表18-2）并画出它们之

间的间隔图象。图18-5就是表18-2矩阵的图象化，

表18—2　　四变量相关矩阵

	A	B	C	D
A	x	0.8	0.2	0.1
B		x	0.5	0.3
C			x	0.9
D				x

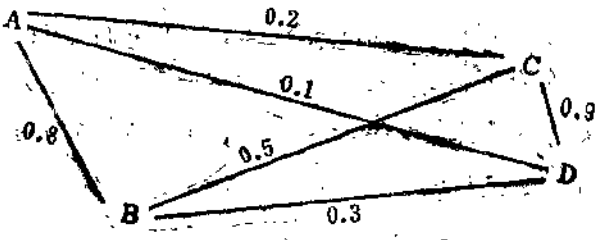


图18—5　最小间隔分析：四变量

图中字母的间隔没有算术意义，只是按相关系数大小加以排列。间隔最远的一对变量（AD）与最弱相关系数相对应；间隔最近的一对变量（CD）与最强的相关系数相对应，余类推。

6 方差分析

方差分析与第十七章讨论的统计显著性检验的

逻辑相同。样本中的个案按自变量的取值分组，然后分析各组在因变量上的区别。方差分析有两种主要形式：单因子方差分析和二因子方差分析。

（1）单因子方差分析

假定我们想比较民主党人和共和党人的收入，看是否共和党人更富裕。我们首先抽样一个样本，了解每个研究对象的党派及收入情况，并分别计算两个组的收入平均数和中位数。结果共和党人的平均年收入为21,000美元，民主党人为19,000美元。显然，共和党人比民主党人富裕。但是这一区别“显著”吗？完全随机地抽取两个组，其平均收入之差是否也会达到2,000美元？

方差分析可用来解决这一问题。数据分布的方差（如收入分布的方差）是度量一组数值离数程度的尺度。图18—6显示了两种可能性。在两种情况下，共和党人的平均收入均为21,000美元，民主党人的平均收入均为19,000美元。但在第一部分，共和党人的收入都集中在21,000美元左右，民主

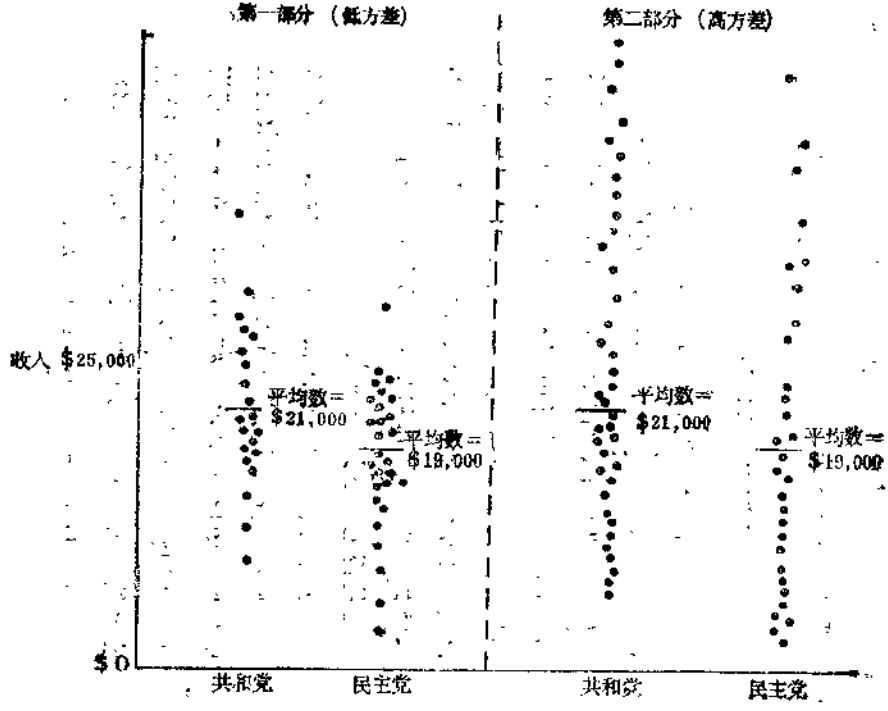


图18—6　共和党人和民主党人收入的两种分布模式

党人集中在19,000美元左右；在第二部分双方收入数据都十分离散。图的第一部分使人相信共和党人同民主党人之间的收入差别确实存在；但第二部分就很难确定：或许差异产生于随机抽样误差？

在方差分析中，是凭统计学的计算而不是一般印象来做判断。经过方差标准化处理的平均值差与数值理论分布的标准统计数表两相比较即可确定平均值差的统计显著性，从而得出关于这一区别显著程度的结论。如，我们可能发现抽样误差只有千分之一，因此可以认为这一区别"在99.9%的水平上是显著的"。

计算过程现在一般都可用计算机进行，最简单的方法之一是两平均值的"t检验"。但如果分组在两个以上，计算就会变得复杂。

（2）二因子方差分析

单因子方差分析适用于双变量分析，如党派和收入。如前所述，社会科学研究常常会涉及多变量分析，二因子方差分析即可用于两个以上变量的检验。如，我们假定民主党人与共和党人收入之差则是由教育程度不同造成的：共和党人比民主党人教育程度高，而教育程度高的人收入就高。二因子方差分析可将两个变量的影响分别开来，与前面讨论

过的阐释模式及部分相关、部分回归的逻辑相似。

7 对数线性分析

假定我们想了解政治倾向与党派倾向的关系。如，是否自由主义者比保守主义者更倾向于加入民主党。首先，把样本分成自由派和保守派两个组，然后分别计算各组中民主党人所占的比例。如发现自由派中民主党人比例高，则可作出政治倾向与党派选择相关这一结论。

在第十四章和第十六章，所有的因变量都只有两个取值。如果因变量不只两个取值，如在民主党人和共和党人之外，样本中还有独立党人、社会党人等等，计算将大大复杂化。如果自变量也不只两个取值，事情就会进一步复杂化。假定在自由派和保守派中间加上一个中间派，比率表就更取做了。

对数线性模式为解决这类复杂问题提供了新方法。它用一种特殊模式来描述变量间的交互作用，将期望值与现察值加以比较（同第十七章卡方检验的逻辑相似）瑞诺德斯（H·T·Reynolds）对这一过程做过如下描述，“象多数统计方法一样，在对数线性分析中，研究者设计一个他认为与数据最

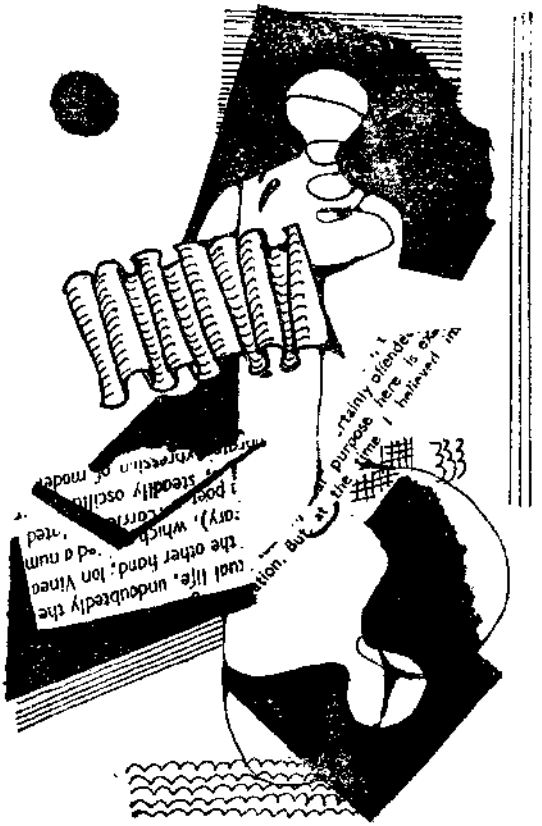
相吻合的模式。这个模式只是一个暂时的关于变量间交互作用的陈述。在选择了这一模式之后，对样本的期望值做出估计，假定这一模式是真的。然后将估计值 $\hat{F}$ 与观察值进行比较。”

在对数线性分析中，研究者要考虑因变量同每一自变量之间的直接关系，每对自变量之间的关系，三变量或多变量之间的关系。假定人们的党派倾向是其政治倾向和种族的函数。这一模式的成分于是有：(1)政治倾向对党派倾向的直接影响；(2)种族对党派倾向的直接影响；(3)种族对政治倾向的影响；(4)种族对政治倾向与党派倾向二者关系的影响；(5)政治倾向对种族与党派倾向二者关系的影响。由于每一成分都有部分的解释力，对数线性分析提供了识别最重要成分及可忽略成分的方法。

对数线性分析在社会科学中的应用日益普遍，但它毕竟是一个新方法。它有两个主要的缺点：第一，某些数据很难满足其数学假设的要求。第二，对数线性分析的结果很难进行直观的简单的比率或平均值比较，因此在可用简单比率表的情况下不应使用对数线性分析方法。它适用于更复杂的情况。

上述只是某些分析方法的简单介绍，还有许多方法尚未提及。本章的目的仅在于引导读者对大量的研究方法作进一步的了解。

第五篇　研究的社会背景



第十九章
社会科学研究中的伦理和政治问题

1 概 论

任何研究工作都会受到四种限制：科学的、实施的、伦理的和政治的。前面的十八章主要集中在科学和实施这两个方面。某些研究设想本身是科学的，但难以实行：或耗资太大或时间要求过长。实际上全书都在讨论如何掌握这二者间的平衡。另外两个重要问题是伦理和政治。有时，研究者出于伦理或政治上的考虑，不得不放弃某些研究项目。

如果某项研究需要对儿童施用体刑，任何人都会立即表示，这项研究是不道德的。如果某人通过访问，了解到人们的性行为，然后把研究对象的话全都发表在地方报纸上，显然也是不道德的。然而，即使是道都高尚的人，也会在某些情况下忽略伦理的问题，不是由于他个人品质不好，而是因为

人人都会犯错误。本章的前半部将讨论社会研究中的伦理问题，即人们公认的道德规范。

政治上的考虑在社会研究中也很微妙、暧昧和充满分歧。本章的后半部将讨论政治对研究的影响。

2 社会研究中的伦理问题

在辞典中，伦理常常与道德并列作为正确与错误的标准，但每个人区分正确与错误的标准各有不同。有的人根据宗教，有的人根据政治意识形态，还有人根据能否办成事情这一实用的原则来作出判断。

韦伯恩特新世界辞典将伦理定义为“符合某个职业或团体的标准”。由于很难定义所谓绝对的道德标准，人们往往视一个人群的约定为日常生活中的道德和伦理。如果你准备在某个特殊的社会里生活，先搞清这个社会的道德规范，定然受益非浅，如果你准备进行社会科学研究，应当首先搞清研究者一般公认的关于适当研究手段的道德规范。本节将对一些社会研究中的重要伦理约定做一综述。

（1）自愿参加的原则

社会研究常常要打扰人们的生活。当访问者按响门铃或一份问卷寄到手中时，就意味着研究对象将开始一种费时、费力而又不是出于自己要求的活动了。参加社会研究打断了研究对象的正常活动。

社会研究还常常要求人们提供那些甚至不愿同朋友同事谈及的私事，而且此类信息还常常会被传达给对研究对象来说是完全陌生的人。某些职业，如医生或律师，也要求此类信息，但基本上是为了满足提供信息者的个人利益。社会研究者却很难以此为理由。象医学科学家一样，社会科学家只能说他们的研究最终会有益于全人类。

在医学研究中，一个主要的伦理原则即参加实验必须是自愿的。这一规范也适用于社会科学研究，不能强迫任何人参加研究。这一规范看似容易做来却难。许多试制中的药物是在囚徒身上进行实验的，虽然参加者是自愿报名的，但许多人的动机是从此类合作中得到好处。社会学教授让学生填写问卷，虽然声明是自愿参加，但学生却担心不参加会影响自己的分数。为了避免此类现象，教授应当做出进一步的表示，如在学生填写问卷时暂时离开教室或请求学生课后填好放在他的信箱里等等。

自愿参加的原则有时会直接与科学研究的要求发生冲突。假如实验对象或填写问卷的人都是自愿参加的，概括性可能会受到威胁，因为自愿参加某项研究这一行为本身可能是某种特殊个性的反映，而实验结果就难以概括各种类型的人了。在一个描述性调查中，只有用科学方法抽出的样本中的大多数人都参加了研究，不论是自愿的还是有点不情愿的，研究者才能把样本的调查结果推论到总体。

由此可知，自愿原则虽然是一个重要的原则，但又常常是难以实行的。当不得不违背这一原则时，重要的事在于保持其他伦理规范，如不伤害研究对象的原则。

（2）对参加者无害的原则

无论参加者是否自愿，社会研究都绝不能伤害研究对象。这个原则在实际中主要是指不能在披露某些信息时使信息提供者感到受窘，或给他们的家庭生活、朋友关系及工作等造成威胁。

此外，社会研究过程还可能使研究对象在心理上受到伤害。如研究对象常常被问到越轨行为、反常表现或使他觉得降低身分如低收入之类的问题。提供此类信息很可能使研究对象感到不快。资料的分析及报告也可能伤害研究对象。当研究对象看到

根据自己参加过的研究写出的书和文章时，可能会认出自己属于其中提到的某一类人。当他们发现自己被归入有偏见的、不爱国的或不信教的类别时，虽然没有指名，他们也会感到某种程度的不安甚至会危害他们的自我想象。

（3）匿名和保密的原则

保护研究对象利益最重要的原则之一是对他们身份的保密。如果发表研究结果会给他们带来任何损害，遵守这一原则就是头等重要的事情。有两项技术可以解决这个问题：匿名和保密，它们常常被混为一谈。

〔匿名〕研究对象应当是匿名的，即研究者不知道研究对象的姓名。具体作法有多种，如邮寄的调查问卷都不编号，问卷寄回之后再加编号。有少数研究对象写上了自己的姓名，应当在见到后立即把它擦掉。

〔保密〕在保密的调查中，研究者知道研究对象的姓名但绝对保密。姓名和地址要尽快从问卷上去掉，代之以编码。可设一单独档案把编码与人名联在一起，以便在必要时与研究对象联系，但除法律需要外绝不向任何人透露。

（4）研究者的身份

有时，研究者对研究对象说明自己的身份不但是有益的而且是必要的。但当你的研究不允许暴露自己的真实身份时，则应慎重考虑伦理规范。欺骗毕竟是不道德的，即使有绝对必要。有时研究者必须隐瞒研究的目的。例如，某公共福利机构要求你调查接受福利待遇人员的生活水平。研究对象可能试图装得比他们的实际情况更穷，因而使调查数据失真。怎么办呢？办法之一是声称为某大学做一般学术研究——隐瞒你与福利机构的关系。这样做可以提高研究的科学质量，但却违反了伦理规范。

（5）分析与报告

社会研究者不仅对研究对象负有伦理的责任，对同行也有这一责任。

首先，研究者最熟悉自己研究的缺点，他有责任在研究报告中指出这些缺点。其次，应当报告否定自己假设的发现和结果，不应只报告肯定的结果。在一项研究中，发现两变量之间并无关系也是很有价值的。在关于士兵对提升制度的意见的研究中，斯托芬坚持要搞清军事警察比空军士兵对提升制度意见少的原因。如果他发现研究结果与预想不

同就放弃分析，那么“参照系”和“相对剥夺”理论的发现就要推迟很久了。

3 两个关于伦理问题的争论

本节举两个引起伦理争论的研究，一个是关于公共厕所中同性恋行为的研究，另一个是关于服从性的研究。

（1）“茶室交易”

研究生兰德·哈姆弗瑞斯（Land Humphreys）对同性恋现象，尤其是一些非同性恋者偶然卷入同性恋事件的行为发生了兴趣。他决定对公园中公共厕所里陌生人之间的同性恋行为进行研究。他的研究结果于1970年发表在一本题为《茶室交易》的书中，因为那些同性恋者戏称公共厕所为“茶室”。

哈姆弗瑞斯特别感兴趣的是，此类活动的参加者看上去在其他方面过着正常的生活，无论在家里还是社区里，没有任何迹象表明他们是同性恋者。因此对他们来说，匿名特别重要。哈姆弗瑞斯发现这种活动一般是三个人进行：两个人干一个人望风，于是他常去“茶室”自愿充当望风的角色，就

这样他实地观察了这种活动。由于不能直接问这些人的姓名地址，他记下了他们的汽车号码，并通过警察查到了他们的姓名地址。然后他对这些人一一进行了化装访问，借口是进行一般调查，从而搞到了在“茶室”搞不到的信息。

哈姆弗瑞斯的研究发表后在学术界内外引起轩然大波。有些人批评他以科学的名义侵犯了人们的私生活。他们认为人们在公共厕所干什么，那是他们自己的事，而不是哈姆弗瑞斯的事。其他的批评集中在欺骗这一点上：他让那些同性恋者以为他和他们是一样的人。批评认为哈姆弗瑞斯追踪到同性恋者家里并化装访问是不道德的。也有些人为哈姆弗瑞斯的研究辩护，认为除非这么做他不可能进行这个研究，而这个题目是值得研究的。虽然他采用了欺骗手段，但没有伤害任何人。

可想而知，关于茶室交易的争论是难以平息的，人们至今仍在辩论，也许会一直辩论下去，因为它触动了感情和伦理问题，伦理问题是难以取得一致意见的。你怎样认为？哈姆弗瑞斯是道德的吗？无论你怎样看，总会有人不同意你的观点。

（2）观察人的服从性

第二个事例与第一个有很多区别。哈姆弗瑞斯

的研究采用参与观察的方法，而这一研究是在实验室中进行的。前述研究属于社会学研究，关于服从性的研究则是心理学研究。

第二次世界大战后，德国士兵惯用的掩饰其暴行的共同借口是："我只是服从命令而已。"如果这个论点正确无异于说不论是谁被放在他们的位置上都会做同样的事。如果上级命令士兵去杀害婴儿，命令本身即可免除士兵对其行为的个人责任。

虽然军事法庭在审判战争罪犯时并不认为这一说法成立，但社会科学家承认在某种程度上这种观点是影响了社会生活的。如果有上级命令，人们常常会做出一般人认为是错误的事情。

为了观察人们在接到命令时能够作出的伤害他人的行为。弥尔格朗（milgram）设计了一个有40个成人参加的实验。实验过程如下：每个研究对象都被告知，他与另一研究对象被分为一组，已通过抽签决定他是"老师"，另一人是"学生"。然后"学生"被带到另一房间绑在椅子上，手腕上捆上电极。"老师"面前有一控制台，上面许多按钮标明从15到315伏特，有些按钮上标明"强电击"、"危险—极强电击"等字样。

实验开始时，"老师"对"学生"读一组词，检验他给这些词配对子的能力。由于"老师"不能

看到“学生”，在“老师”的控制台上有一个信号灯显示“学生”的答案。每当答案出错时，实验主持人就要求“老师”按一个按钮给“学生”一个电击——从最轻的开始，逐渐加重。两个房间中间的门是敞开的，因此“老师”可以听到“学生”对电击的反应。随着实验的进行，“老师”给“学生”的电击越来越强，直到“学生”尖叫着哀求“老师”发发慈悲停止实验。但实验主持人坚持实验必须继续下去。又过了一段时间，“学生”开始踢墙和大声嘶叫。实验主持人仍坚持让“老师”继续实验。最后，当“老师”又读了一组新词时，“学生”那边无声无息，室内一片沉寂。实验主持人对“老师”说，没有回答等于答错了，并命令“老师”给“学生”一个最强的电击。

如果你是“老师”，会在实验进行的哪一阶段拒绝继续给“学生”电击呢？当“学生”第一次尖叫起来的时候？当他用脚踢墙的时候？还是当他全无声息的时候？在参加弥尔格朗实验的40人中，在学生踢墙之前拒绝继续实验的人一个也没有，到踢墙时有5个人拒绝继续实验下去，而三分之二即26个人一直做完了一切要求他们做的事，包括最强的电击。

你可能猜到了，电击是假的，“学生”是由另

一个实验人员扮装的。只有"老师"是这个实验中的真正研究对象。他并没有伤害任何人，尽管他以为伤害了。这个实验的目的在于研究人的服从性。

弥尔格朗的实验在方法上和伦理上全都受到攻击。伦理方面的批评集中在实验对研究对象的影响。有些人对电击有过个人经验，因此在实验过程中请求停止，他们极端愤怒和紧张不安，有的甚至暴跳如雷。

你对这个研究有什么看法？你是否认为这个研究题目十分重要以致可以采用这种研究手段？你能否想出其他办法来研究人的服从性？

4 社会研究中的政治问题

人们在政治上的分歧象伦理上的分歧一样多，但二者又有区别。首先，虽然伦理和政治常常交错在一起，但社会研究中的伦理问题一般仅涉及研究所使用的方法，政治问题往往涉及研究的实际内容和应用。以弥尔格朗实验为例，伦理方面的批评认为，他使用的方法对研究对象有伤害；政治方面的批评则会认为服从性是一个不应当研究的课题。其中一种观点会认为根本不应当试图改变人们服从上

级命令的愿望；另一种观点则会认为这项研究的结果会使人们变得更加盲目服从。

其次，在伦理上有公认的规范，但在政治上却没有。固然，有些伦理规范也有其政治的一面（如：从保护公民权的角度来说，研究不应伤害研究对象），但没有一整套搞社会研究公认的政治规范。唯一的例外是一般认为研究者个人的政治倾向不应当影响科学研究。人们可以有自己的政治观点，但进入科学领域时应当把它放在一边。在这个意义上，科学被理想化为非政治性的、非道德性的和绝对客观的。

（1）客观性与意识形态

第一章曾谈到社会研究不可能是绝对客观的，因为研究者有个人主观性。科学作为一个集体的事业是通过许多主观性达到客观性。换言之，虽然各个科学家具有各不相同的主观看法，但如果采用的研究方法是可以站得住的，就能够而且应当得到相同的结果。问题的提法应当是看人们能把个人的价值观放弃到什么程度。

关于社会科学的客观性和中立性的问题，经典的论述是马克斯·韦伯在1918年的一个题为“作为事业的科学”的讲演中提出的。他在演说中提出

“超价值观的社会学”，认为社会学要想对社会作出独特的贡献，就要象其他科学一样清除个人的价值观。自由派和保守派却应当承认“事实”，无论从他们各自的政治立场上来看这一事实是好的还是坏的。

多数社会科学家已经在这个抽象的理想上取得了一致。近年来，马克思主义和新马克思主义学者们指出，社会科学和社会行动不能、也不应当被截然分开。他们认为，对社会地位的解释很微妙地起到为社会地位辩护的作用；对种族歧视一类偏见的社会功能的解释很容易使其继续存在合法化。同时，仅仅研究社会及其弊病而不想改造它，就是不负责任。

除能否有“超价值观”的社会科学这一抽象分歧之外，对于某些具体研究是超价值观的还是表现了研究者本人的政治价值观，也有很多争论。

试举一例。再没有什么研究领域象种族关系的研究那样与政治难解难分了。社会科学家研究这个题目已有很长历史，其研究成果往往与现实政治纠缠在一起。

当美国最高法院于1896年确立“种族隔离但平等”这一原则时，虽然并未请教也没引证任何社会科学的研究，但最高法院受到威廉·格瑞汉·萨莫

（William Graham Summer）著作的影响却是公认的事实。作为那个时期主要的社会科学家之一，萨莫有一著名的观点，认为一个社会的风俗习惯比起立法来说是更难改变的。他反复论证过“国家政策不能创造风俗习惯”这一论点。因此，最高法院否定了“用立法来克服社会偏见”的假设以及“使法律与社会的一般观念发生冲突”的想法。

当“种族隔离但平等”的原则在1954年被推翻时，最高法院的新决定是在考虑到种族隔离对黑人儿童的有害影响这一论点后作出的。为了证实这一论点，最高法院引证了一批社会学和心理学的研究报告。

总的说来，社会科学家在本世纪一直是支持美国的种族平等事业的。许多人成为公民权运动的积极分子。有些人甚至更加激进。因此，将支持种族平等观点的社会科学家不必担心同行们的批评。但也有少数研究得到与占统治地位的意识形态立场不一致的结论。

1966年，当詹姆·科利曼（Jame Coleman），一个受尊敬的社会学家发表了一项关于种族和教育的研究结果时，引起了一场热烈的争论。与一般公认的看法相反，科利曼发现黑人学生的学习成绩在种族隔离或种族混合的学校没有显著区别，家庭和

居住区域的因素才对学习成绩产生重大影响。

许多在公民权运动中相当活跃的社会科学家都不愿接受科利曼的研究成果。有些人从方法论上批评他的著作，更多的人激烈地反对他，因是认为他的发现会带来种族隔离主义的政治后果。

举这个例子旨在说明政治意识形态常常会卷入社会研究。虽然科学的抽象标准是脱离意识形态的，但科学的实践却很难脱离意识形态。

（2）科学与政治

政治及其意识形态的作用并非只存在于社会科学领域，自然科学也曾经遇到而且继续面对同样的情况。本书将科学与政治的关系作如下概括：

首先，科学并不能完全脱离政治，尤其是社会科学，它是社会生活的一部分。我们研究的事物是人们所关心的事物，是人们对之有个人感受的事物，是会影响他们的生活的事物。科学家也是人，而其作为人的感觉常常会通过他们的所作所为表现出来。不承认这一点是幼稚的。

其次，即使是在政治争论和相互仇恨的气氛之中，科学研究还是会继续进行的。尽管研究者互相攻讦或承受着来自学术界之外的攻击，科学研究仍

在继续进行，报告在继续发表，新的事物逐步为人们所理解。简言之，意识形态的争议不但不会阻止科学的进程，反而会使它变得更加激动人心。

最后，意识形态的考虑应当成为我们创造性工作的部分背景，它可以使我们更加自觉地去学习多种多样的社会科学的方法和技巧。许多科学方法可以帮助我们发现和改正缺点，特别是那些我们尚未察觉的缺点，否则我们就会在这个世界上除了自己——自己的偏见和信仰——之外，什么也看不到。

走向未来丛书

一九八四年出版

人的发现	李平晔著
增长的极限 （罗马俱乐部关于人类困境的研究报告）	李宝恒译
激动人心的年代	李醒民著
GEB——一条永恒的金带	道·霍夫斯塔特原著 乐秀成编译
现代物理学与东方神秘主义	根据F·卡普拉的《物理学之道》编译　澹　耕编译
现实与选择	朱嘉明　吕　政著
经济控制论	何维凌　邓英淘编著
冒险与世界	于有彬编著
看不见的手	杨君昌编著
语言学与现代科学	陈明远编著
在历史的表象背后	金观涛著
让科学的光芒照亮自己	刘青峰著

走向未来丛书

一九八五年出版

人的现代化 〔美〕阿历克斯·英格尔斯等著 殷陆君编译
大变动时代的建设者 汪家溶编著
没有极限的增长 朱利安·林肯·西蒙原著
黄江南 朱嘉明编译
西方社会结构的演变 金观涛 唐若昕著
在国际舞台上 陈汉文编著
昨天 今天 明天 邓正来编著
摇篮与墓地 陈越光 陈小雅著
择优分配原理 茅于轼著
第三次浪学危机 胡作立著
凯恩斯革命 杨君昌编著
艺术魅力的探寻 林兴宅编著
西方文官系统
杨百揆 陈子明 陈兆钢 李盛平 廖晓非著
动态经济系统的调节与变化 邓英淘 何维凌编著
斯的魅会 〔美〕爱德华·奥尔本·威尔逊著
李昆峰编译

走向未来丛书

一九八六年出版

富饶的贫困	王小强 白南风著
定量社会学	郭治安 姜 璐 沈小峰编译
儒家文化的困境	蒋功秦著
系统思想	〔美〕小拉尔夫·弗·迈尔斯主编 杨志信 葛明浩译
日本为什么“成功”	〔日〕森岛通夫著 胡国成译
困壮的衰落	金观涛 王翠衔著
郭洛伊德著作选	约翰·里克曼编 贺明明译
西方的丑学	刘 东著
十七世纪英西的科学、技术与社会	〔美〕R·K·默顿著 范岱年 吴 忠 蒋效东译
困布上的制造	戴士和著
德启超与中西近代思德	〔美〕约瑟夫·阿·勒文森著 刘 伟 刘 丽 姜铁军译
新教伦困与资本主义境神	〔德〕马克斯·韦伯著 黄晓京 彭 强译
困想革命的技术困流	宋德生著
增长、短缺与效率	〔匈〕亚诺什·科尔内原著 崔之元 钱铭今译

走向未来丛书

一九八七年出版

走向现代国家之路　钱乘旦 陈意新著
竞争中的合作　陈汉文编著
计圈历史学　〔苏〕科瓦尔琴科主编　闻　一 肖　吟译
哲学的还原　麦克斯韦·约翰·查尔斯沃斯著　田晓春译
凯恩斯现论与中国经济　林　一知著
人的创世纪　张　猛 顾　昕 张继宗编著
社会研究方法　〔美〕艾尔·巴比著　李银河译
发展社会学　胡格韦尔特著　白　桦 丁一凡编译
上帝怎样掷骰子　陈克艰著
空寂的神殿　谢选骏著
震撼心灵的古旋律　郑　凡著
以权力制约权力　朱光磊著
整体的哲学　金观涛著
人体文化　谢　长 葛　岩著
人心中的历史　刘　昶著
渐进渐的模式　罗首初 万解秋著
发展的主题　周其仁 杜　鹰 邱继成著
社会选择与个人价值　〔美〕肯尼思　阿罗著
陈志武 崔之元译
对科学的猜想与漫见　〔英〕查　帕　斯诺著
陈恒六 刘　兵译
马克斯·韦伯　〔英〕弗兰克·帕金著　刘　东 谢维和译